

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: July 30, 2004, 14:16:25 ; Search time 3640 Seconds
(without alignments)
7251.623 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725
Perfect score: 609
Sequence: 1 atgataagaatcaatgacct.....cggataaccattatcaaaacta 609

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 3470272 seqs, 21571516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: *
1: gb.ba: *
2: gb.htg: *
3: gb.in: *
4: gb.om: *
5: gb.ov: *
6: gb.pat: *
7: gb.ph: *
8: gb.pl: *
9: gb.pr: *
10: gb.ro: *
11: gb.sts: *
12: gb.sy: *
13: gb.un: *
14: gb.vi: *
15: em.ba: *
16: em.fun: *
17: em.hum: *
18: em.in: *
19: em.mu: *
20: em.om: *
21: em.or: *
22: em.ov: *
23: em.pat: *
24: em.ph: *
25: em.pl: *
26: em.ro: *
27: em.sts: *
28: em.un: *
29: em.vi: *
30: em.htg.hum: *
31: em.htg.inv: *
32: em.htg.other: *
33: em.htg.mus: *
34: em.htg.pln: *
35: em.htg.rod: *
36: em.htg.mam: *
37: em.htg.vrt: *
38: em.sy: *
39: em.htgo.hum: *
40: em.htgo.mus: *
41: em.htgo.other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	21	3.4	110000	8	AC145127_20	Continuation (21 o
2	21	3.4	153510	8	AC099734	AC099734 Oryza sat
3	21	3.4	303511	8	AE017053	AE017053 Oryza sat
4	21	3.4	313050	1	BX321857	BX321857 Nitrosomo
5	20	3.3	2080	10	MMU011080	AJ011080 Mus muscu
6	20	3.3	22190	9	AC117443	AC117443 Homo sapi
7	20	3.3	86385	2	AC138385	AC138385 Mus muscu
8	20	3.3	86385	2	AC138385	AC138385 Mus muscu
9	20	3.3	121201	9	AC128685	AC128685 Homo sapi
10	20	3.3	154934	2	AC098177	AC098177 Homo sapi
11	20	3.3	157793	9	AC090181	AC090181 Homo sapi
12	20	3.3	171061	2	AC109266	AC109266 Mus muscu
13	20	3.3	175066	9	AC005900	AC005900 Homo sapi
14	20	3.3	180172	2	AC115834	AC115834 Mus muscu
15	20	3.3	184180	9	HS1042K10	AL022238 Human DNA
16	20	3.3	221618	2	AC098411	AC098411 Rattus no
17	19	3.1	315	10	MUSZEP92	DI0631 Mus muscu
18	19	3.1	855	6	AR317790	AR317790 Sequence
19	19	3.1	2202	9	BC034912	BC034912 Homo sapi
20	19	3.1	3582	10	BC053927	BC053927 Mus muscu
21	19	3.1	32436	1	AB101202	AB101202 Acinetoba
22	19	3.1	45643	3	U23177	U23177 Caenorhabdi
23	19	3.1	64201	2	AC117833	AC117833 Homo sapi
24	19	3.1	64900	2	AC121510	AC121510 Mus muscu
25	19	3.1	67339	2	AC138375	AC138375 Mus muscu
26	19	3.1	68930	9	AC068136	AC068136 Homo sapi
27	19	3.1	82165	9	AC117471	AC117471 Homo sapi
28	19	3.1	85397	3	AC026301	AC026301 Caenorhab
29	19	3.1	90024	2	AC139690	AC139690 Homo sapi
30	19	3.1	97137	9	AC004745	AC004745 Homo sapi
31	19	3.1	110000	9	AE014305_0	AE014305 Homo sapi
32	19	3.1	118401	9	AC091305	AC091305 Homo sapi
33	19	3.1	129245	9	AL359547	AL359547 Human DNA
34	19	3.1	133652	9	AC010326	AC010326 Homo sapi
35	19	3.1	142624	2	AC128628	AC128628 Rattus no
36	19	3.1	152631	9	AC117432	AC117432 Homo sapi
37	19	3.1	155656	9	EX293560	EX293560 Human DNA
38	19	3.1	155757	2	AC079984	AC079984 Homo sapi
39	19	3.1	156678	2	AC145026	AC145026 Sus scrof
40	19	3.1	158770	9	AC007321	AC007321 Homo sapi
41	19	3.1	162475	9	AC146163	AC146163 Pan trogl
42	19	3.1	163288	2	AC119264	AC119264 Mus muscu
43	19	3.1	164061	9	AL445188	AL445188 Human DNA
44	19	3.1	164394	9	AC116904	AC116904 Homo sapi
45	19	3.1	167398	9	CNS072ET	AL445886 Human chr

ALIGNMENTS

RESULT 1
AC145127_20
WPCOMMENT
Sequence split into 24 fragments LOCUS AC145127 Accession AC145127

Fragment Name	Begin	End
AC145127_00	1	110000
AC145127_01	100001	210000
AC145127_02	200001	310000
AC145127_03	300001	410000
AC145127_04	400001	510000
AC145127_05	500001	610000
AC145127_06	600001	710000
AC145127_07	700001	810000
AC145127_08	800001	910000
AC145127_09	900001	1010000


```
/note="Similar to Oryza sativa gene, repeat sequence
Micropan-1"
5518..5539
/note="Simple Sequence Repeat (AT)n"
5542..5575
/note="Simple Sequence Repeat (AT)n"
5602..5627
/note="Simple Sequence Repeat (AT)n"
5724..5745
/note="Simple Sequence Repeat (AT)n"
5868..6457
/note="Simple Sequence Repeat (TA)n"
8249..15154
/gene="OSJNBb0004A06.3"
/note="Unknown protein"
join(8249..8785,9076..9120,9930..10065,11299..11387,
11908..11950,12602..12789,13220..13423,14642..15154)
/gene="OSJNBb0004A06.3"
/codon_start=1
/product="Unknown protein"
/protein_id="AA111187.1"
/db_xref="GI:22953954"
/translation="MNNLVNKKGLTDHANNKQKKEKGRNSDWHGSGTKDTIPFM
GDADNADGVVGVACGTOAPALLLLCPHOOLMOIQSKCCYVTSASSTGGR
GQAWAMVIAWRPADKGTAADEDEGWLPIARQGRDQKGRSRRPRAERQ
APFVVELASRRRDGSPVTRSSYSLLITGQALNQBYRKYSAWTERSLIGS
SLARIOESCSGRNRNGKSIKFLIKOLFVDESHLEFLNDRFLPCEAPKRSK
TKQNLQSSIKCNKTIWGGISLHQLNQDFVDFRGRKNYIIVILRRSG
IICGALKIEPKPIFPHRSFPVPLRIENGDSMTTCHQGVNSGVIKTPFH
ICKTSVPQFQIITLLIGSPSEAFIDMEESKNKSIKVEFRCHFSPQFNQEQ
QKQGRSGTSQSLMPLDKWAKSEKALTLGAGTSGVQPPSQAVHTTASCKEKKDK
EPNMQSTPEVGTREMLVHSGTKVKKLVSSITIEDHVSNSKNVTVGKVMPEIQSVP
LVHVSBSASQNRFT"
9056..9073
/note="Simple Sequence Repeat (AT)n"
10560..10589
/note="Simple Sequence Repeat (AT)n"
10975..11020
/note="Simple Sequence Repeat (AT)n"
complement(19177..22551)
/gene="OSJNBb0004A06.4"
/note="Contains similarity to disease resistance protein"
complement(join(19377..21480,22331..22551))
/gene="OSJNBb0004A06.4"
/codon_start=1
/product="Putative disease resistance protein"
/protein_id="AA111188.1"
/db_xref="GI:22953955"
/translation="MGRGTEGRSSGAGSGHPAPVAGKEREVEVASGLPTLSLPE
EVERKNGRRRGKVRKARQKMGGRIFLDVDDISAKTKWTKISYLLTNKE
RSRIIVSRQVAGSCCPENKDLPIPIPRDSKELFNRSVSEKSTDRHKVQ
NNVFDLWRCGGQPLAVTWAGLVACNQPQKSYWAGLHKLPEEVSITAGAQOET
SLNLDGVTRILDCNNLPGYRTCLLYLAIFPKGKISKCLSRWIAEGFVNAKQ
LTABEASVFNHILRRKLIRPVHSSNGKLTFOVHDMVLDIVTKAREENFTVVG
GHLMMIAPNNKVRSLSQSSSKHSGDSTKGNLSQVRSITVFGSLTPPEAFNDRI
LOVDFQSLGKFKEMHCKMKVLAFLSRTDTHVPTIVKLEYLETLDRETR
VKELPVEVEQKLSRILGSKSNKPNRKLRLPQKSKQKHSMUTQKKEGMKAL
RILSGIKDITVAGHVLGTLKLTLYKLNPEEPGRKLTIRSSIEYLSGCG
LQTLAINEGSNFINSLGNMSPRYVALSGLMKLPPGKWTIRLTSLKITSLS
VTRDTLEHALPLSLTFSFGEMEKDQKMKDIIEDNKSLSGDSIFPVGEGFK
SLKLRLFPVLPVPLGFCNMAPLEMIQFOAFELGIDTLGNLKGVRLEAKPR
BEDQTAQIDLLVRLDKDSTEGKVIIDHTFTS"
complement(24383..25405)
/gene="OSJNBb0004A06.5"
/note="Contains similarity to disease resistance protein"
complement(24383..25405)
/gene="OSJNBb0004A06.5"
/codon_start=1
/product="Putative disease resistance protein"
/protein_id="AA111189.1"
/db_xref="GI:22953956"
/translation="MELVAGSTATWDSILGKLNLLAGEYDLIRVGRGDIQYISDEL
ASMAKAFLLDLAREDPNRKKNWKQIRDMAYDCEDCIDDFAHRLPNDLSLAKCCPWV
```

```
TLVYDLMTFGPREIASSIAELKVAQLIADRRIRYGVNPNWTKQKGPDPATSDIA
EDQLASHELGNFPGMEKAKDLREWDTACQFAVYVIGVFGKGTITAMALYK
KMVYQFCRAWTVSQNDLDVINDILKIDPDYQCSSKTGTSENITLARFGSK
LKEDVQRTGSLROSSPRESIBETSNLKRITETTDNKLESQIKLLDKRCVHPHSIS
ISISY"
repeat_region 26913..27062
/note="Similar to Tourist Oll MITE element"
complement(27190..27338)
repeat_region
Query Match 3.4%; Score 21; DB 8; Length 153510;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 340 CAAGTACTTTCATGCTGTGGT 360
|||||
Db 110715 CAAGTACTTTCATGCTGTGGT 110735
|||||
RESULT 3
AE017053 303511 bp DNA linear PLN 06-JUN-2003
LOCUS Oryza sativa (japonica cultivar-group) chromosome 10, section 7 of
DEFINITION 77 of the complete sequence.
ACCESSION AE017053 AE016959
VERSION AE017053.1 GI:31430010
KEYWORDS Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 303511)
AUTHORS The Rice Chromosome 10 Sequencing Consortium
CONSTRM In-depth view of structure, activity, and evolution of rice
TITLE Chromosome 10
JOURNAL Science 300, 1566-1569 (2003)
REFERENCE 2 (bases 1 to 303511)
AUTHORS Buell, C.R., Wang, R.A., McCombie, W.R., Messing, J. and Yuan, Q.
DIRECT Direct Submission
TITLE Submitted (05-MAY-2003) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr., Rockville, MD 20850, USA
COMMENT This is the pseudomolecule for rice chromosome 10, which was
constructed by resolving discrepancies between overlapping BACs,
trimming the overlap regions, and linking the unique sequences to
form a contiguous sequence. Genes in individual BAC clone were
identified by a combination of several methods: Gene prediction
programs, searches of the complete sequence against a peptide
database and EST databases. Genes with similarity to other proteins
are named after the database hits. Genes without significant
peptide similarity but with EST similarity are named as unknown
proteins. Genes without protein or EST similarity, that are
predicted by more than two gene prediction programs over most of
their length are annotated as hypothetical proteins. Genes
encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES
location/Qualifiers
source 1. 303511
organism="Oryza sativa (japonica cultivar-group)"
mol_type="genomic DNA"
cultivar="Nipponbare"
db_xref="taxon:39947"
chromosome="10"
misc_feature <1..79862
/note="Chromosome Sequence Derivation: nucleotide sequence
in this region was derived from BAC clone OSJNBb0072F04
(GB:AC092553)."
gene complement(<6121..>7701)
locus_tag="OSJNBb0072F04.10"
/note="Contains similarity to NPK1 related protein kinase"
complement(<6121..>7701)
mRNA /locus_tag="OSJNBb0072F04.10"
CDS complement(6121..7701)
```

```
/locus_tag="OSJNBb0072F04.10"
/codon_start=1
/product="putative NPX1-related protein kinase"
/protein_id="AAP51982.1"
/db_xref="GI:31430011"
/translation="MAAIIIGGGWTRLSVRGASGAVSLAANDVSGELFVKSAG
EGARQQLREWSVMSGLSSPHVLCGLFVQAAAGAGGSEQLLEAVPGSLA
RNGRLDEGATVYADVLGLDYLHGLVGHGVSNGVLVAGDGAKEAKLTDGFCARV
ANPGSKQVLGGTAFMAPEVARGEEOGLAADVWALGCTVIEMATGRAPWSDNVL
PALHKIGYDAVDPLPRLMSPEAKDFLGRCLQRAGDRPTAAQLQHPPTSKSCGLN
KRTVKATVSPSALDALTMSSESSDTGDEEDVMSNSPTGRIRAMACSCQTLPDW
SDHGCSWLEVLGSVINAKTAIEORVTSMACSPSSVPDMSQGNQGLDVLSSV
ISIANLETRATRADVSSECPAKWVRMACSPSSVPDMSQGNQGLDVLSSV
EDVAAAQGISQAAVSVIGVSGSEQSVVVENQEDFETSLSCSERVLVGVHAAD
NNAASKAKIKCSNFSC"
complement (<10528..>12105)
/notes="contains similarity to NPX1 related protein kinase"
/locus_tag="OSJNBb0072F04.11"
/codon_start=1
/product="putative NPX1-related protein kinase"
/protein_id="AAP51983.1"
/db_xref="GI:31430012"
/translation="MAAATIGGGGWTLSRIGHGASGATVSLAADAAGSLFVYKS
ACDAVATARQQLREWSVMSGLSSPHVLCGLFVQAAAGAGGSEQLLEAVPGSLA
DVVARNGRDLSSAFRAYADVRLGLDYLHSEKLVVGHGVKSNVLVAGDRAKLADEG
CARVMPGSKQFVLGGTDFAFVAPVARGESQGPAAVWALGCTVIEMATGRAPWSDM
DVLALRMIGYDTPDLPPWLSPKDFLRCWQRAGDRPTAAQLQHPFVSKSC
GLKEVKTATVPTSDALDAALAESETSSDDEADNNSPTGRIIAMAASGQC
TLPDMSDHWGIEVLGTVSINAKKTAADAEASBPAPRAMACSPSSVPDMS
SDNHGIDVLSPADDDGECQNAPEEDVVAADQIFGEAVGSIIVGVSGSEQSVVVE
NORDEPISVYSCSERILVAHVAADNNAASKAKIKESCHDPRPSIPSRCAHNLFLSL
IFEQIDNSNVPPAPKY"
complement (<16098..>17177)
/notes="unknown protein"
/locus_tag="OSJNBb0072F04.12"
/codon_start=1
/product="unknown protein"
/protein_id="AAP51984.1"
/db_xref="GI:31430013"
/translation="MAKPTSLPRLVIGARWDAERPLGRLLILAHAAFLDAGFVPA
AARADNSIRLPKRGRTASSLPYAAPQLLHPDDAAVQLRICALHRLVLYVNA
RCSMFMDLTVKVCIDLAALALGGALDDTARALRDARLAALWALADRLCRRVL
VDVCAENGVTLPPTFMSLPDDVKAAILARLPDGDLDARAECTCAGLRRLVADRDRDAA
LAKPRYKLPFLQLIGSDDDGGPTTEVSKKKYVAARLWPGELFASMRTRPLP
IYAPLLDDEFSETFWFDKPSLPPEIIVPFRRRRRWRAMPDAGHGLAARGH
GDKKPRHAGAVHSPSRFRWKHR"
complement (<19076..>20170)
/notes="unknown protein"
/locus_tag="OSJNBb0072F04.13"
/codon_start=1
/product="unknown protein"
/protein_id="AAP51985.1"
/db_xref="GI:31430014"
/translation="WADTSLPRLHRIICAAWDAEPLGLRLVILAHAAFLDAGFVSTGAA
NIDGACQSVLPQVATASLSRYTAPQLLHRRDOPAAAAAATAVALRVCAGRRH
VYVYCVRFANPLDTYICIDAPAAALAGGLDDTARALRERCARGLAAWSLA
DLRCRVLLCAKNGVPEPEHELMSPDDVKVAILARLAAGADLDLAVCTVCGLNL
DLAHDSTLWKMYPYKLSQRLRRLRFLVSGYBPTAVSKAAVAVRERVAADVD
FWGEILPLPWTSMRVPMIRRYTPFPFPFPSPPEEETVPRRRRRAMPDAGHR
AAPGHGDKQWRGAGVHSPSSGFRWKHR"
complement (<25322..>29537)
/locus_tag="OSJNBb0072F04.14"
/notes="contains similarity to copia type pol polyprotein"
complement (join (<25322..>26796,27148..>27678,27734..>27814,
28019..>28279,28375..>28542,29526..>29537))
/locus_tag="OSJNBb0072F04.15"
complement (join (<25322..>26796,27148..>27678,27724..>27814,
28019..>28279,28375..>28542,29526..>29537))
/locus_tag="OSJNBb0072F04.16"
/codon_start=1
/product="putative copia-type pol polyprotein"
/protein_id="AAP51986.1"
/db_xref="GI:31430015"
/translation="MSYKALVAKENVMIVDSCVSRHMGDKNWFSLKASKTSRLI
FEDASTSVLAIGLVKVKCKTKGSKVDFSCGDSVLIIISYGRVFXADFPNVPVITC
LAAPFDXDMFWERRLGVDFDLRLSLGSDQGVQGVKVKGLNLDLDEAFCHERGLFLR
LDLFFPGSLKRIEFDNGSLSEHSPRPQDNGVVERKVKLVLEAMARMDDEYKTPKR
FWAEINATCYISNRFLSKLGTSEYLFQGVKSHLVLSCKFCVILKSGNDLDF
EARTDGLFGLYPAHSGRYLVLETKIIEETCEVDFEASPTGRIIAGTISVOGGE
DGRIFDESDNNDNDESGAELENFERNKVTVPFEPSEHNFPGTKWFMQNDQSDGL
IVRNKTRIVAQGFQVEGLDDETFAHVARVEAIRLLAFATSKGFKLVQMDKGFDFL
NGPTQBEVYKQPPENPFPHVKLSKALYGLQAPRAMVDREKFLAKGSEMG
KIDKILFVLKRGDNLVQIYVDDLIIFGSHALVVEPAETMREREMESMGLSEYEL
GLQIKTPQGFVHOMKITDILLRRFKNCKRPSITPIGTAVLDDDECEAVDOKEY
RSMIGSLYLIAGRLDIQALCLYTRPPASPRASHRAQIKRMKRLNHLTFEFGIWTST
SSISLGSYSDTFGGCRIDRKTSGICHPLETSLIASSRKQSSVAQTSSESVAA
ASCCSQDWLWLSLTKYKSLTFEKVPLFCNTSAINIAKNPVORSHKTHINICFHLRD
HVEKGDELQFLDKMLQLADIENKPLDSNHFALRGELGIITHPPGVV"
complement (<32428..>42847)
/locus_tag="OSJNBb0072F04.15"
/notes="contains similarity to transposase"
join (<32428..>32464,33104..>33314,33384..>33512,33627..>33798,
33937..>34750,34784..>35140,36883..>36958,38204..>38231,
39651..>39752,40093..>40296,40509..>40808,41038..>41846,
42146..>42316,42395..>42695,42788..>42847)
/locus_tag="OSJNBb0072F04.15"
join (<32428..>32464,33104..>33314,33384..>33512,33627..>33798,
33937..>34750,34784..>35140,36883..>36958,38204..>38231,
39651..>39752,40093..>40296,40509..>40808,41038..>41846,
42146..>42316,42395..>42695,42788..>42847)
/locus_tag="OSJNBb0072F04.15"
/codon_start=1
/product="putative transposase"
/protein_id="AAP51987.1"
/db_xref="GI:31430016"
/translation="MVKDKTVPVNGWPKQGNIDIVYFKPLVEDLHLLTWGVQWMD
TYKHEHFCBALIGTITDYPAIGNTSGOMVCKMCAVFLARSHPHONKXAFDQTV
EIDRPIFRGSSILAEVKLHERRARGIVEHTWQDEGHAFSTERSHKWTSGKCIL
KTEMAATTYLRHAILLVQKKLLPVALRIGLEKIRDFIKLCCOFFKAINKQVIDP
ASLGMQEDTILNLCQDMLPFLSFDIMVHLVHLVEQVKILGTAYLDOSMAFRFWM
GRIIKYVNRISIPEGSIIEGYATEVEVEFIDYMDIQLPIGVPPREHGMKGLIGH
RRIIVAPNRQIRIAHFAVLSQALVAPYMDMLDLVLRNDRPKGAWLAKMHKLOFGK
WLMVYLTADIDGAHFSLSWPSNTIHMVQDMNGYTFVTSQDNKSINQNSGRVID
AYDESVEGLDYGALVPLPQCKVKVIRGLKIDKHGTIIIVLDNDVDEDEPVLAKS
VKQIPYKDSLYKMLKPHVLEDKOSIVGMEISTKMTTINMPSHRPQLMSLKTROA
LAKMTKASAAVKGSGVRRGGGSGNDGPMPSRCVYEVGTEHKEFASSEGEHH
ASPTFNLNQHSEEGEBCWRKXLSFTFPEGBELAKGALSKMGLFREWTDL
NVDFVQCNRTDPFEKMYQKLADYDFEGSTIGSGELVALGHQHELKEKITEAISL
ARQGFVYVREKDEITVALGTKEHGMRGVSPWPKGFPDDTMYRARSPTREE
EAGRMLEVAYQSNITOEIARNVAHPICAGQTHGVLPADHSRVEVEVPSQYCYV
LOFPLEBGGTTLAALHNLVWPKXIVITNDGAGSDSDSSGNSGRHSPRSRSP
SLPQKTPNQPOPSQKSPSTGIRKKTTPSASQMSGSSSMERLPHSKSLVNPVPTT
SLPKYSGSITSLIGPKKPERKIPKVIIEHFKCAQPLPPIPKPIIDFRRTLEKSE
MQSRRLFLDKKASDEAAEKSVLDKSLDDINDLKDRAVLSNMDDPMRWKSYOCRS
LDKRIAFDPAVNVNQLSKEREIDYFNALVKGNDHILLPILSHHHLVFIN
IDDSKICVYDSLRKGTNYQTIMNALRAYKRRSKRTYGRKTCADATASFRIFENQYI
YRQALNLCGMVYMYMLCFVSGHLLPRNAEKLGLETSEMLPHVFTALTD"
complement (<48333..>52859)
/locus_tag="OSJNBb0072F04.16"
/notes="contains similarity to NBS LRR type resistance
protein"
complement (join (<48333..>48348,48499..>48586,49139..>52859))
/locus_tag="OSJNBb0072F04.16"
complement (join (<48333..>48348,48499..>48586,49139..>52859))
```



```
/locus_tag="OSJNB0072F04.16"

Query Match      3.4%; Score 21; DB 8; Length 303511;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 340 CAAGTACTCAGCTGGTGGT 360
|||||
DB 218168 CAAGTACTCAGCTGGTGGT 218188

RESULT 4
BX321857/c
LOCUS Nitrosomonas europaea ATCC 19718, complete genome, segment 2/10.
DEFINITION BX321857 AL954747
ACCESSION BX321857.1 GI:30138174
VERSION complete genome.
KEYWORDS Nitrosomonas europaea ATCC 19718
SOURCE Nitrosomonas europaea ATCC 19718
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
          Nitrosomonadaceae; Nitrosomonas.
REFERENCE 1
AUTHORS Chain, P., Lamerdin, J., Larimer, F., Regala, W., Land, M., Hauser, L.,
          Hooper, A., Klotz, M., Norton, J., Sayavedra-Soto, L., Arciero, D.,
          Holmes, N., Whittaker, M. and Arp, D.
TITLE Complete genome sequence of the ammonia-oxidizing bacterium and
          obligate chemolithoautotroph Nitrosomonas europaea
JOURNAL J. Bacteriol. 185 (9), 2759-2773 (2003)
MEDLINE 22586410
PUBMED 12700235
REFERENCE 2 (bases 1 to 313050)
AUTHORS Larimer, F.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2002) Submitted on behalf of the Nitrosomonas
          genome consortium, the DOE Joint Genome Institute, Production
          Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
          USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
          1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
          larimerf@ornl.gov
REMARK Nitrosomonas genome consortium
FEATURES
    source
        1..313050
            location/Qualifiers
                /organism="Nitrosomonas europaea ATCC 19718"
                /mol_type="genomic DNA"
                /strain="ATCC 19718"
                /db_xref="taxon:228410"
                complement(132..503)
                /locus_tag="NE0270"
                complement(132..503)
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAD84181.1"
                /db_xref="GI:30138175"
                /translation="MPEQITRKASPFCLRLTPERTLLERAAGLPLGEVIRQOVFE
                NRKRSRKQPKVDHRLLSGLGELGRSLANNQLARAACGLLNLTPVEKTSLL
                NACADLRHRTLMKSLGNS"
                complement(385..1253)
                /locus_tag="NE0271"
                complement(385..1253)
                /locus_tag="NE0271"
                /function="InterPro IPR002559"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="transposase IS4 family"
                /protein_id="CAD84182.1"
                /db_xref="GI:30138176"
                /translation="MFLSLSPQAGDAPEGRKLLKLENKGFSDTHVIMDKAYEGDETR
                QLVLDLGIMPVPPKANRVPWEYDKRNEVERLFRRLKFRFRIFSRFDKLDVS

misc_feature
    complement(887..1698)
    /note="ISNE5"
    /locus_tag="NE0272"
    /complement(1253..1660)
    /locus_tag="NE0272"
    /function="COGS COG3293"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="possible transposase"
    /protein_id="CAD84183.1"
    /db_xref="GI:30138177"
    /translation="MEISASQFKLIENLPIQRGNTVLNVLNAILVYAEKGCWR
    GLPVKGNWHSIYTRNWARNGVLDRVFLAQCNKLIQLEVDHMSLDSTIVKHPDG
    TGALKTKTVKLSVNHGAGLPKFTWSRQMPER"
    complement(1853..2488)
    /gene="cyc"
    /locus_tag="NE0273"
    complement(1853..2488)
    /gene="cyc"
    /locus_tag="NE0273"
    /function="InterPro IPR000345:IPR002329:IPR003098 COGS
    COG2863"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="Cytochrome c, class I; Cytochrome c, class I"
    /protein_id="CAD84184.1"
    /db_xref="GI:30138178"
    /translation="MKFIVTVAIALTFSGVFABSPAGDVEKKEIAAGICACGH
    NADGNSAIPYPILAGQYPGYIAKQNDPKVEGETVKRDNQIMAPMVAITLQDDLEN
    LAAYSQKQPGTASDASLVETGKLYQGNLENSIPACSSCHSPNGGQIPHPYRI
    DQHPAYTLSQLAFRQGTAKNTNTMTQITVSRMSEQEQAVSEYIATLR"
    /gene="cyc"
    /locus_tag="NE0273"
    /note="Signal predicted by SignalP 2.0 HMM (Signal peptide
    probability 1.000) with cleavage site probability 0.999 at
    residue 22"
    2595..3197
    /locus_tag="NE0274"
    2595..3197
    /locus_tag="NE0274"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAD84185.1"
    /db_xref="GI:30138179"
    /translation="MRKFSVFLLANIFVVFYLHGRPDNLPAQIALIHSEKIELLPA
    KVACLKWNELGVPVOHVHVEISKWESGQDHTEISRGVTVHWHIPLRNARETAK
    QIEQLKKSGLSYLHIQENADSPWENAIISLALPDDSDVAALVELKKGKGVRLMDSEQ
    VLEQFEFDIRNPTEQITESVRQLAQQFPETKLEVTCSR"
    2595..2660
    /locus_tag="NE0274"
    /note="Signal predicted by SignalP 2.0 HMM (Signal peptide
    probability 0.892) with cleavage site probability 0.890 at
    residue 22"
    3311..4129
    /gene="map"
    /locus_tag="NE0275"
    3311..4129
    /locus_tag="NE0275"
    /function="InterPro IPR000994:IPR001714:IPR002467 COGS
    COG0024"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="Methionine aminopeptidase"
```

/protein_id="CAD84186.1"
 /db_xref="GI:30138180"
 /translation="MQAGAGVVIHIKTAPEIETMTIAGRLASEVLDVIEPVVPGVTT
 GELDELCHRYMDVQYTIIPAPLYAPPYSPKSICTSVNHQVCHGVGPKLKDGD
 VNLDTIVYEGVHGTSYMYVGFPSIQAKRLCELTVEAMWRIETIRPGKHLGDIG
 HAJORLAGSVSVREFCGHIGAKFHEDPOVLHYGRAGTGIELKPCMFTVPEPMIN
 AGKAAIKQLPDQWTYITKDHLSLAQWHEHTILVDSYEVLTIVSSGSPARPAYR"
 4197..4922
 /gene="rph"
 /locus_tag="NE0276"
 4197..4922
 /gene="rph"
 /locus_tag="NE0276"
 /EC_number="2.7.7.56"
 /function="interPro IPR001247:IPR002381 COGS COG0689"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="3' exoribonuclease family"
 /protein_id="CAD84187.1"
 /db_xref="GI:30138181"
 /translation="MPCNNRAPAQMRPVRIIRHYVRAHGSVLIEYGETRVICTASV
 IEKVPPLKAGAGWLTARYGMLPRSTGCRQREAAKGKQSRIMEQLRIGRLRSI
 LDLEKGETIQYDQCDVTCADGTRTASITGAFVADYDAIDYLAERMIQNPIDRV
 AVSVGLKQGLLEDDYLEDSDGCDTDLNVMTGSLGLVEVQGTAEKRVVFSRQELVY
 LNAVQGLQLEFDVQRKALETVA"
 4960..5571
 /locus_tag="NE0277"
 4960..5571
 /locus_tag="NE0277"
 /function="interPro IPR002637 COGS COG0127"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="Ham1 family"
 /protein_id="CAD84188.1"
 /db_xref="GI:30138182"
 /translation="MNKIVIASNNAGKLAETSRLLAPLGIEVWTQSSILGVTEADPHM
 TFVENALAKARHASLATGDPALADSGICVSRALGDPGVFSRAGVPRSDERNRKL
 VEALHGDPRRAYVYCVILVPHGQDQPVLIETDTCETIETAEPIGOGGYPHPFL
 PELGKTAAELSEENRSHRGKALRLVQLSENETVPVVPV"
 5608..6807
 /locus_tag="NE0278"
 5608..6807
 /locus_tag="NE0278"
 /function="interPro IPR003401:IPR004559 COGS COG0635"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="Coproporphyrinogen III oxidase and related Fes
 oxidoreductases"
 /protein_id="CAD84189.1"
 /db_xref="GI:30138183"
 /translation="WPVLSPDPSKPSHLAELPPLGLYLHPWCVRKPCYCDENSHAIN
 SGDLDPESVYVALKEDIELALPLVGRPVTVSPGCGTSPLESATADAILSHVRML
 LRLAPGVEITLLEANGTLEAOKFTDFHAGLNRLSIGIQSNRHLQALGRIDGAE
 LRAEALQHFNLDMYALQPDTRDEALQDIETACNRGVHAISHUTIEPTFF
 YVFPQLPDEDSADQMVIEMHILAEGRYHYETSAFQPGKPLHNNWYQVGYDIF
 IGAGHSKISFRDRIIRQNRHHPQYLOEAMAGNACLENLLATEQVGPQDRIFPM
 MNALRTIEGDFPLEFTRTGLTITIASIQSLAEQRLIEWQHDICRPTPEGRFFLND
 LLEIFLS"
 6979..8820
 /gene="pcoa"
 /locus_tag="NE0279"
 6979..8820
 /gene="pcoa"

Query Match 3.4%; Score 21; DB 1; Length 313050;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

427 GCCCGGGTACTGCTCAAAAT 447

Db 121608 GCCCGGGTACTGCTCAAAAT 121588

RESULT 5

MMU011080/c 2080 bp mRNA linear ROD 24-FEB-1999
 LOCUS Mus musculus mRNA for alpha-albumin protein.
 DEFINITION Mus musculus mRNA for alpha-albumin protein.
 ACCESSION AF011080
 VERSION A011080.1 GI:3646360
 KEYWORDS alpha-albumin.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1. Van Reeth, T., Gabant, P., Dreze, P., Szpirer, J. and Szpirer, C.
 Unpublished
 JOURNAL 2 (bases 1 to 2080)
 AUTHORS Van Reeth, T.
 TITLE Direct Submission
 JOURNAL Submitted (15-SEP-1998) Van Reeth T., departement de Biologie
 moleculaire, Universite Libre de Bruxelles, laboratoire de Biologie
 du Developpement, Rue des Chevaux, 57, 1640 Rhode St Genese,
 BELGIUM

FEATURES
 Location/Qualifiers
 1..2080
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /chromosome="5"
 /clone="IMAGE 522099"
 /census_type="diaphragm"
 /clone_lib="Stratagene mouse diaphragm (#937303)"
 /dev_stage="adult"
 1..2080
 /gene="alpha-albumin"
 72..1907
 /gene="alpha-albumin"
 /codon_start=1
 /product="alpha-albumin protein"
 /protein_id="CAA09471.1"
 /db_xref="GI:3646361"
 /db_xref="GOA:O89020"
 /translation="MRHLKLTGIFPILLPTLESALPTKPDQVDHFNATQKIDENTA
 YLIAIATSOYVQASDFEVLTKVMDYDRCDWADNTLPECKSTANDAIQMLCDME
 GLPKHNFSCGKAGPRELCFYNNKANVGFPPETLDPBEKQAVKNSGFLH
 LYVEARRNPFVPVLLAVAFSEAAATCCQOQKATCFQAKAPITQYLKASS
 YQRYVCALIKFGPKVNLNSINAVFSKPKFQKLTLLLEDVSMYSGCGGVVH
 CIRQSVNHICSKQDSISSKICVCEKTLREACTINANKDRPEGLSREKAFV
 ESENVCDSDPDKFFABFIYISRRHDLSTPRLIRITKVMDFLEDCSRENPA
 CYRVEDKFNETQRSLAVQOCNQFELGKTLQRFHVKFTKAAPQLPMEELVSL
 SKENVAALITCCILDEFACVNLADLVGLCGVNTNRTINPAVDHCKTDFARRH
 CFELKADITVALPSVSALVSLHTDWCOPKXEDLQKXHLFVNLVKMMPGITDBEW
 LCLTKTKTAARESCSEVQEPESCFSPESSKGTGDSQATENKITDQ"

ORIGIN
 Query Match 3.3%; Score 20; DB 10; Length 2080;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 CCAGTTTGAACCTCTCTGG 401
 Db 1864 CCAGTTTGAACCTCTCTGG 1845

RESULT 6

AC117443/c

LOCUS

DEFINITION

BAC Library) complete sequence.

AC117443

Accession

22190 bp DNA linear PRI 26-SEP-2002

Hom sapiens 3 BAC RP11-208N8 (Roswell Park Cancer Institute Human

BAC Library) complete sequence.

AC117443

AC117443.4	GI:22024339
HTG.	Homo sapiens (human)
KEYWORDS	Homo sapiens
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ORGANISM	1 (bases 1 to 22190)
REFERENCE	Mundy D.M., Adams C.C., Adio-Obiola B., Ali-osman F.R., Allen C., Albrooks S.L., Amarutunge H.C., Are J.R., Ayele M., Banks T., Barbaria J., Benton J., Bmaga K., Blankenburg K., Bonin D., Bouck J., Bowie S., Brieva M., Brown E., Brown M., Bryant N.P., Bunay C., Burch P., Burke K., Burrell K.L., Byrd N.C., Charon T.F., Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C., Davey-Carroll L., Dederich D.A., Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J., Earnhart C.B., Edgar B., Edwards C.C., Elhaj C., Escotto M., Falls P., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R., Gilletti J.H., Guevara W., Gunaratne P., Hale S., Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A., He X., Hernandez J., Hernandez O., Hodgson A., Hogues M., Holloway C., Hollins B., Homsi P., Howard S., Huber J., Hulky S., Hume J., Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah J., Kovach S., Kratovic J., Kureshi A., Landry N., Leal B., Lewis D.C., Lewis D., Li J., Lichkarge O., Lieu C., Liu J., Liu W., Louisgied H., Lozado R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J., Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E., Massey E., Mayhew E., McLeod M.P., Meador M., Mei G., Metzger M., Miner G., Miner Z., Mitchell T., Mohabbat K., Moore S., Morgan M., Moorish T., Morris S., Moser M., Neal D., Nelson D., Newton J., Newton N., Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokwenkwu S., Ogih M., Okunishi G., Oragunde N., Pickens R., Pace A., Payton B., Peery J., Perez L., Peters L., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savory G., Scherer S., Scott G., Shen H., Shoohtart N., Siesson I., Sodergren E., Sonaike T., Sparks A., Stanley H., Stone H., Sutton A., Syatek A., Taber P., Tanerisa A., Tanerisa K., Tang H., Tansey J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S., Usmani K., Vasquez L., Vera V., Villalon B., Vinson R., Wang Q., Wang S., Ward-Moore S., Warren R., Washington C., Watlington S., Williams G., Williamson A., Wleczyk R., Woodson S., Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Naylor S.L., Weinstein G. and Gibbs R.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 22190)
AUTHORS	Worley K.C.
TITLE	Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL	3 (bases 1 to 22190)
REFERENCE	Worley K.C.
AUTHORS	Direct Submission
TITLE	Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL	4 (bases 1 to 22190)
REFERENCE	Worley K.C.
AUTHORS	Submitted (31-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL	5 (bases 1 to 22190)
REFERENCE	Worley K.C.
AUTHORS	Submitted (26-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL	

Baylor Plaza, Houston, TX 77030, USA On Jul 31, 2002 this sequence version replaced gi:20335979. INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu	CLONE LENGTH: This sequence does not necessarily represent the entire length of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.
ANNOTATION OF FEATURES: STRs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts. Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished), for Human and Mouse sequences. Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.	SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.
QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.	Location/Qualifiers 1..22190 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="3" /clone="RP11-208N8" complement(1..2005) /note="overlaps bases 1..2005 of clone AC024106"
misc_feature	/function="clone overlap" 259..292 /rpt_family="AT_rich"
repeat_region	1130..1225 /rpt_family="MIR"
repeat_region	complement(1326..1495) /rpt_family="MLTIA2"
repeat_region	complement(2527..2598) /rpt_family="MIR"
repeat_region	2843..2871 /rpt_family="AT_rich"
repeat_region	2911..2933 /rpt_family="AT_rich"
repeat_region	3435..3468 /rpt_family="CT-rich"
repeat_region	complement(4422..5459) /rpt_family="L1PAL2"
repeat_region	5586..5607 /rpt_family="(TA)n"
repeat_region	6549..6580 /rpt_family="(TAAA)n"
repeat_region	7165..7195 /rpt_family="AT_rich"
repeat_region	7920..8035 /rpt_family="(TA)n"

```

repeat_region      complement(8414..8473)
                    /rpt_family="HERVL"
repeat_region      8474..8796
                    /rpt_family="MSR44A"
repeat_region      complement(8797..8913)
                    /rpt_family="HERVL"
repeat_region      complement(8919..9763)
                    /rpt_family="HERVL"
repeat_region      10824..11174
                    /rpt_family="THE1B"
repeat_region      complement(11511..11925)
                    /rpt_family="MLT1B"
repeat_region      12102..12131
                    /rpt_family="(CTA)n"
repeat_region      complement(13658..13691)
                    /rpt_family="U6"
repeat_region      14610..14747
                    /rpt_family="MLT1I"
repeat_region      14827..15888
                    /rpt_family="L2"
repeat_region      15963..16008
                    /rpt_family="AT_rich"
repeat_region      16053..16120
                    /rpt_family="AT_rich"
repeat_region      16373..16475
                    /rpt_family="(TATATG)n"
repeat_region      complement(16816..17006)
                    /rpt_family="MER83C"
repeat_region      complement(17179..17300)
                    /rpt_family="MER83-internal"
repeat_region      17432..17489
                    /rpt_family="(TG)n"
repeat_region      complement(18943..19012)
                    /rpt_family="HUERS-P3"
repeat_region      complement(19013..19367)
                    /rpt_family="THE1B"
repeat_region      complement(19368..19835)
                    /rpt_family="HUERS-P3"
misc_feature       complement(20204..22190)
                    /note="overlaps bases 94715..96701 of clone AC128685"
                    /function="clone overlap"
repeat_region      complement(20701..21077)
                    /rpt_family="MER66-internal"
repeat_region      complement(21804..22190)
                    /rpt_family="MER83-internal"

```

ORIGIN

```

Query Match      3.3%; Score 20; DB 9; Length 22190;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 322 AATGGCCTGAATTCCTCA 341

Db 21719 AATGGCCTGAATTCCTCA 21700

```

RESULT 7
AC138385
LOCUS      AC138385      86385 bp      DNA      linear      HTG 27-DEC-2002
DEFINITION Mus musculus clone RP24-383E18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC138385
VERSION   AC138385.1 GI:27375071
KEYWORDS HTG; HTGS PHASE0.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 86385)
          Birren,B., Nuebaum,C. and Lander,B.
          Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          (bases 1 to 86385)
TITLE     Mus musculus, clone RP24-383E18
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 86385)
          Birren,B., Nuebaum,C., Lander,B., Allen,N., Anderson,S.,

```

```

Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kellis,C., Landers,I., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meidrim,J., Mcneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tsifaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zaincun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26666
Center clone name: 383_E_18
-----
* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 972: contig of 972 bp in length
* 973 1072: gap of 100 bp
* 1073 2033: contig of 961 bp in length
* 2034 2133: gap of 100 bp
* 3005 3104: contig of 871 bp in length
* 3105 4089: contig of 985 bp in length
* 4090 4189: gap of 100 bp
* 4190 5167: contig of 978 bp in length
* 5168 5267: gap of 100 bp
* 5268 6264: contig of 997 bp in length
* 6265 6365: gap of 100 bp
* 7350 7449: gap of 100 bp
* 7450 8407: contig of 958 bp in length
* 8408 8507: gap of 100 bp
* 8508 9478: contig of 971 bp in length
* 9479 9578: gap of 100 bp
* 9579 10554: contig of 976 bp in length
* 10555 10654: gap of 100 bp
* 10655 11636: contig of 982 bp in length
* 11637 11736: gap of 100 bp
* 11737 12693: contig of 957 bp in length
* 12694 12793: gap of 100 bp
* 12794 13729: contig of 936 bp in length
* 13730 13829: gap of 100 bp
* 13830 14828: contig of 999 bp in length
* 14829 14928: gap of 100 bp
* 14929 15916: contig of 988 bp in length
* 15917 16016: gap of 100 bp

```

TITLE
JOURNAL

COMMENT

16017 17012: contig of 996 bp in length
17013 17112: gap of 100 bp
17113 18087: contig of 975 bp in length
18088 18187: gap of 100 bp
18188 19181: contig of 994 bp in length
19182 19281: gap of 100 bp
19282 20257: contig of 976 bp in length
20258 20357: gap of 100 bp
20358 21402: contig of 1045 bp in length
21403 21502: gap of 100 bp
21503 22492: contig of 990 bp in length
22493 22592: gap of 100 bp
22593 23578: contig of 986 bp in length
23579 23678: gap of 100 bp
23679 24649: contig of 971 bp in length
24650 24749: gap of 100 bp
24750 25707: contig of 958 bp in length
25708 26799: gap of 100 bp
26799 26899: gap of 100 bp
26900 27894: contig of 995 bp in length
27895 27994: gap of 100 bp
27995 29003: contig of 1009 bp in length
29004 29103: gap of 100 bp
29104 30120: contig of 1017 bp in length
30121 31193: contig of 973 bp in length
31194 31293: gap of 100 bp
31294 32268: contig of 975 bp in length
32269 32368: gap of 100 bp
32369 33353: contig of 985 bp in length
33354 33453: gap of 100 bp
33454 34379: contig of 928 bp in length
34380 34479: gap of 100 bp
34480 35464: contig of 985 bp in length
35465 35564: gap of 100 bp
35565 36537: contig of 973 bp in length
36538 36637: gap of 100 bp
36638 37628: contig of 951 bp in length
37629 37728: gap of 100 bp
37729 38721: contig of 993 bp in length
38722 38821: gap of 100 bp
38822 39817: contig of 996 bp in length
39818 39917: gap of 100 bp
39919 40919: contig of 1002 bp in length
40920 41019: gap of 100 bp
41020 41969: contig of 950 bp in length
41970 42069: gap of 100 bp
42070 43057: contig of 988 bp in length
43058 43157: gap of 100 bp
43158 44163: contig of 1006 bp in length
44164 45276: gap of 100 bp
45277 45376: gap of 100 bp
45377 46363: contig of 987 bp in length
46364 46463: gap of 100 bp
46464 47423: contig of 960 bp in length
47424 48459: gap of 100 bp
48460 48559: contig of 936 bp in length
48560 49553: gap of 100 bp
49554 49653: gap of 100 bp
49654 50640: contig of 987 bp in length
50641 50740: gap of 100 bp
50741 51735: contig of 995 bp in length
51736 51835: gap of 100 bp
51836 52840: contig of 1005 bp in length
52841 52940: gap of 100 bp
52941 53926: contig of 996 bp in length
53927 54026: gap of 100 bp
54027 55022: contig of 996 bp in length
55023 55122: gap of 100 bp
55123 56111: contig of 989 bp in length

* 56112 56211: gap of 100 bp
56212 57188: contig of 977 bp in length
57189 58253: contig of 965 bp in length
58254 59353: gap of 100 bp
59354 59663: contig of 910 bp in length
59664 60378: contig of 1015 bp in length
60379 60478: gap of 100 bp
60479 61471: contig of 993 bp in length
61472 61571: gap of 100 bp
61572 62570: contig of 999 bp in length
62571 62670: gap of 100 bp
62671 63657: contig of 987 bp in length
63658 63757: gap of 100 bp
63758 64741: contig of 984 bp in length
64742 64841: gap of 100 bp
64842 65799: contig of 958 bp in length
65800 65899: gap of 100 bp
65900 66904: contig of 1005 bp in length
66905 67004: gap of 100 bp
67005 67983: contig of 979 bp in length
67984 68083: gap of 100 bp
68084 69072: contig of 989 bp in length
69073 69172: gap of 100 bp
69173 70169: contig of 997 bp in length
70170 70269: gap of 100 bp
70270 71267: contig of 998 bp in length
71268 71367: gap of 100 bp
71368 72375: contig of 1008 bp in length
72376 72475: gap of 100 bp
72476 73471: contig of 996 bp in length
73472 73571: gap of 100 bp
73572 74574: contig of 1003 bp in length
74575 75670: contig of 996 bp in length
75671 75770: gap of 100 bp
75771 76750: contig of 980 bp in length
76751 76850: gap of 100 bp

Query Match 3.3%; Score 20; DB 2; Length 86385;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 ACTTGTCTCACTGCTCTTGG 143

Db 8746 ACTTGTCTCACTGCTCTTGG 8765

RESULT 8

AC138385/c

LOCUS

DEFINITION

AC138385

AC138385

AC138385.1

HTG; HTGS PHASE0.

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.

1 (Bases 1 to 86385)

Authors

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus, clone RP24-383E18

Unpublished

2 (Bases 1 to 86385)

Authors

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Bloom, T., Boguslavsky, I., Boukhgaltier, E.,

Canarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,

Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,

Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,

Gardyna, S., Gord, S., Graham, J., Grand-Pierre, N., Hafez, N.,

Hagos, B., Horton, I., Hulme, M., Illiev, J., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,

AC138385 86385 bp DNA linear HTG 27-DEC-2002
Mus musculus clone RP24-383E18, LOW-PASS SEQUENCE SAMPLING.

AC138385

AC138385

AC138385.1

HTG; HTGS PHASE0.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.

1 (Bases 1 to 86385)

Authors

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus, clone RP24-383E18

Unpublished

2 (Bases 1 to 86385)

Authors

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Bloom, T., Boguslavsky, I., Boukhgaltier, E.,

Canarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,

Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,

Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,

Gardyna, S., Gord, S., Graham, J., Grand-Pierre, N., Hafez, N.,

Hagos, B., Horton, I., Hulme, M., Illiev, J., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,

Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,
 Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T.,
 Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (27-DEC-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submission@genome.wi.mit.edu
 ----- Project Information
 Center project name: L26666
 Center clone name: 383_E_18

NOTE: This record contains 80 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 * 973 1072: contig of 972 bp in length
 * 1073 1072: gap of 100 bp
 * 2034 2033: contig of 961 bp in length
 * 2134 2133: gap of 100 bp
 * 3005 3004: contig of 871 bp in length
 * 3105 3104: gap of 100 bp
 * 4090 4089: contig of 985 bp in length
 * 4190 4189: gap of 100 bp
 * 5168 5167: contig of 978 bp in length
 * 5268 5267: gap of 100 bp
 * 6265 6264: contig of 997 bp in length
 * 6365 6364: gap of 100 bp
 * 7350 7349: contig of 985 bp in length
 * 7450 7449: gap of 100 bp
 * 8408 8407: contig of 958 bp in length
 * 8508 8507: gap of 100 bp
 * 9479 9478: contig of 971 bp in length
 * 9579 9578: gap of 100 bp
 * 10555 10554: contig of 976 bp in length
 * 10655 10654: gap of 100 bp
 * 11637 11636: contig of 982 bp in length
 * 11737 11736: gap of 100 bp
 * 12694 12693: contig of 957 bp in length
 * 12794 12793: gap of 100 bp
 * 13730 13729: contig of 936 bp in length
 * 13830 13829: gap of 100 bp
 * 14829 14828: contig of 999 bp in length
 * 14929 14928: gap of 100 bp
 * 15917 15916: contig of 988 bp in length
 * 16017 16016: gap of 100 bp
 * 17013 17012: contig of 996 bp in length
 * 17113 17112: gap of 100 bp
 * 18088 18087: contig of 975 bp in length
 * 18188 18187: gap of 100 bp
 * 19281 19280: contig of 994 bp in length
 * 19282 19281: gap of 100 bp
 * 20257 20256: contig of 976 bp in length
 * 20258 20357: gap of 100 bp
 * 20358 21402: contig of 1045 bp in length
 * 21403 21502: gap of 100 bp
 * 21503 21492: contig of 990 bp in length
 * 22493 22592: gap of 100 bp
 * 22593 22578: contig of 988 bp in length
 * 23579 23678: gap of 100 bp
 * 23679 24649: contig of 971 bp in length
 * 24650 24749: gap of 100 bp
 * 24750 25707: contig of 958 bp in length
 * 25708 25807: gap of 100 bp
 * 25809 25799: contig of 992 bp in length
 * 26800 26899: gap of 100 bp
 * 26900 27994: contig of 995 bp in length
 * 27995 29003: contig of 1009 bp in length
 * 29004 30103: gap of 100 bp
 * 30104 30120: contig of 1017 bp in length
 * 30121 30220: gap of 100 bp
 * 30221 31193: contig of 973 bp in length
 * 31194 31293: gap of 100 bp
 * 31294 32268: contig of 975 bp in length
 * 32269 33353: gap of 100 bp
 * 33354 33453: gap of 100 bp
 * 33454 34379: contig of 926 bp in length
 * 34380 34479: gap of 100 bp
 * 34480 35464: contig of 985 bp in length
 * 35465 35564: gap of 100 bp
 * 35565 36537: contig of 973 bp in length
 * 36538 36638: gap of 100 bp
 * 36639 37628: contig of 991 bp in length
 * 37629 37728: gap of 100 bp
 * 37729 38721: contig of 993 bp in length
 * 38722 38821: gap of 100 bp
 * 38822 39817: contig of 996 bp in length
 * 39818 39917: gap of 100 bp
 * 39919 40913: contig of 1002 bp in length
 * 40920 41019: gap of 100 bp
 * 41020 41969: contig of 950 bp in length
 * 41970 42069: gap of 100 bp
 * 42070 43057: contig of 988 bp in length
 * 43058 43157: gap of 100 bp
 * 43158 44163: contig of 1006 bp in length
 * 44164 44263: gap of 100 bp
 * 44265 45276: contig of 1013 bp in length
 * 45277 45376: gap of 100 bp
 * 45377 46363: contig of 987 bp in length
 * 46364 46463: gap of 100 bp
 * 46464 47423: contig of 960 bp in length
 * 47424 47523: gap of 100 bp
 * 47524 48459: contig of 936 bp in length
 * 48460 48559: gap of 100 bp
 * 48560 49553: contig of 994 bp in length
 * 49554 49653: gap of 100 bp
 * 49654 50640: contig of 987 bp in length
 * 50641 50740: gap of 100 bp
 * 50741 51735: contig of 995 bp in length
 * 51736 51835: gap of 100 bp
 * 51836 52840: contig of 1005 bp in length
 * 52841 52940: gap of 100 bp
 * 52941 53928: contig of 986 bp in length
 * 53929 54028: gap of 100 bp
 * 54029 55022: contig of 996 bp in length
 * 55023 55122: gap of 100 bp
 * 55123 56111: contig of 989 bp in length
 * 56112 56211: gap of 100 bp
 * 56212 57188: contig of 977 bp in length
 * 57189 57288: gap of 100 bp
 * 57289 58253: contig of 965 bp in length
 * 58254 59263: gap of 100 bp
 * 59264 59363: gap of 100 bp

```

* 59364 60378: contig of 1015 bp in length
* 60379 60478: gap of 100 bp
* 60479 61471: contig of 993 bp in length
* 61472 61571: gap of 100 bp
* 61572 62570: contig of 999 bp in length
* 62571 62670: gap of 100 bp
* 62671 63657: contig of 987 bp in length
* 63658 63757: gap of 100 bp
* 63758 64741: contig of 984 bp in length
* 64742 64841: gap of 100 bp
* 64842 65799: contig of 958 bp in length
* 65800 65899: gap of 100 bp
* 65900 66904: contig of 1005 bp in length
* 66905 67004: gap of 100 bp
* 67005 67983: contig of 979 bp in length
* 67984 68083: gap of 100 bp
* 68084 69172: contig of 989 bp in length
* 69173 70169: gap of 100 bp
* 70170 70269: contig of 997 bp in length
* 70270 71267: contig of 998 bp in length
* 71268 71367: gap of 100 bp
* 71368 72375: contig of 1008 bp in length
* 72376 72475: gap of 100 bp
* 72476 73471: contig of 996 bp in length
* 73472 73571: gap of 100 bp
* 73572 74574: contig of 1003 bp in length
* 74575 74674: gap of 100 bp
* 74675 75670: contig of 996 bp in length
* 75671 75770: gap of 100 bp
* 75771 76750: contig of 980 bp in length
* 76751 76850: gap of 100 bp

Query Match
Best Local Similarity 3.3%; Score 20; DB 2; Length 86385;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 ACATTGCTCAACTGCTTGG 143
Db 80518 ACATTGCTCAACTGCTTGG 80499

RESULT 9
AC128685
LOCUS
DEFINITION
Homo sapiens 3 BAC RP11-535124 (Roswell Park Cancer Institute Human
BAC library) complete sequence.
AC128685
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM

AC128685 121201 bp DNA linear PRI 21-JUN-2003
Homo sapiens 3 BAC RP11-535124 (Roswell Park Cancer Institute Human
BAC library) complete sequence.
AC128685
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 121201)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,W., Brown,B., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,X.J., Draper,K., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gall,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Hawlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsa,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

```

```

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peary,J., Perez,I., Peters,L.,
Pickens,R., Primus,E., Pu,i.L., Quiles,M., Ren,i., Rivers,M.,
Rojas,A., Rojuckan,i., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E.,
Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 121201)
Worley,K.C.
Direct Submission
Submitted (22-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 121201)
Worley,K.C.
Direct Submission
Submitted (13-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 121201)
Worley,K.C.
Direct Submission
Submitted (25-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 121201)
Worley,K.C.
Direct Submission
Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 121201)
Worley,K.C.
Direct Submission
Submitted (21-JUN-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 21, 2003 this sequence version replaced gi:23307944.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
SNPs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST

```


(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht>

FEATURES

```

source
1..121201
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-535124"
misc_feature
1..2094
/notes="overlaps bases 176428..178521 of clone AC112644"
/function="clone overlap"
repeat_region
1..499
/rpt_family="LIMA3"
510..723
/rpt_family="LIMA3"
735..2212
/rpt_family="LIMA3"
2212..3980
/rpt_family="LIMA3"
3981..4016
/rpt_family="(TTTA)n"
complement(4017..4168)
/rpt_family="LIPAB"
4169..4349
/rpt_family="LIMA3"
4691..4720
/rpt_family="LIMA3"
4756..4789
/rpt_family="AT_rich"
5854..5902
/rpt_family="AT_rich"
6406..6462
/rpt_family="AT_rich"
complement(7193..8279)
/rpt_family="MER52A"
complement(8280..8524)
/rpt_family="LIPB1"
8525..8571
/rpt_family="(TA)n"
complement(8672..10086)
/rpt_family="LIPB1"
10087..10516
/rpt_family="LIPB1"
complement(10517..11662)
/rpt_family="LIPB1"
11659..11801
/rpt_family="LIMA4"
12308..12628
/rpt_family="THE1B"
12629..12693
/rpt_family="THE1B-int"
12802..13984
/rpt_family="THE1B-int"
14730..15081

```

```

/rpt_family="THE1B-int"
15082..15435
/rpt_family="THE1B"
15971..16501
/rpt_family="LIMA4"
16514..16799
/rpt_family="LIMA4"
complement(16812..17269)
/rpt_family="LIPAB"
17270..17614
/rpt_family="LIMA4"
18192..18274
/rpt_family="MER41B"
18720..18761
/rpt_family="(TG)n"
complement(18849..19480)
/rpt_family="LIPB3"
20294..20335
/rpt_family="AT_rich"
20506..20802
/rpt_family="Tigger1"
21358..21407
/rpt_family="AT_rich"
23313..23495
/rpt_family="MIRB"

```

Query Match 3.3%; Score 20; DB 9; Length 121201;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 AATGGCCTGAATTCCTCA 341

Db 95186 AATGGCCTGAATTCCTCA 95205

RESULT 10

AC090177/c

LOCUS

DEFINITION

AC090177.3

VERSION

AC090177.3

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 154934)

AUTHORS

Birren, B., Nusbaum, C. and Lander, E.

TITLE

Homo sapiens chromosome 15, clone RP11-762M18

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 154934)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Boguski, L., Boukhalil, B., Brown, A.,

Camarata, J., Campopiano, A., Choquet, Y., Colangelo, M., Collins, S.,

Collins, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S.,

Dodge, S., Faros, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,

Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliiev, I., Johnson, R.,

Jones, C., Karatas, A., Lacombe, K., Lamazares, R., Landers, T.,

Lehoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,

Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,

McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V.,

Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,

O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,

Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Ratta, R.,

Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,

Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,

Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,

Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,

Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zaincun, J.,

Zembek, L., Zimmer, A. and Zody, M.

Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE AUTHORS

3 (bases 1 to 157793)
Birken, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barina, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArallano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, V., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (04-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE AUTHORS

4 (bases 1 to 157793)
Birken, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barina, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArallano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, V., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (17-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE AUTHORS

5 (bases 1 to 157793)
Birken, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barina, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cooke, A., Cooke, P., Corum, B., DeArallano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasany, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (18-MAR-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 17, 2003 this sequence version replaced gi:23308099.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12167
Center clone name: 797_A_18

FEATURES

Location/Qualifiers	source
1..157793	
/organism="Homo sapiens"	
/mol_type="Genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="15"	
/map="15"	
/clone="RP11-797A18"	
/clone_lib="RPC1-11 Human Male BAC"	
/complement(384..433)	
/rpt_family="L1MA2"	
/complement(434..728)	
/rpt_family="AluYc2"	
459..469	
/note="<30 qual SNGL region"	
626..632	
/note="<30 qual SNGL region"	
/complement(729..1439)	
/rpt_family="L1MA2"	
2613..2618	
/note="<30 qual SNGL region"	
2622..2910	
/rpt_family="AluJo"	
2676..2680	
/note="<30 qual SNGL region"	
2713..2717	
/note="<30 qual SNGL region"	
2918..3209	
/rpt_family="AluSx"	
/complement(6825..7135)	
/rpt_family="AluY"	
7645..7993	
/rpt_family="MLTII"	
/complement(8466..8864)	
/rpt_family="L1MC1"	
/complement(8869..9073)	
/rpt_family="L1MB3"	
/complement(9074..9365)	
/rpt_family="AluSx"	
/complement(9366..9376)	
/rpt_family="L1MB3"	
/complement(9377..9680)	
/rpt_family="AluSx"	
/complement(9681..10002)	
/rpt_family="L1MB3"	
10003..10023	
/rpt_family="(TA)n"	
/complement(10024..10310)	
/rpt_family="AluYc"	
/complement(10311..10538)	
/rpt_family="L1MB3"	
/complement(10538..10573)	
/rpt_family="L1MC1"	
11743..12053	
/rpt_family="AluJo"	
12054..12161	
/rpt_family="L2"	
12582..13759	
/rpt_family="L1MB7"	

Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome

Mon Aug 2 09:36:50 2004

us-09-868-338-7_copy_1117_1725.oligo.rge

```

DEFINITION Homo sapiens chromosome 17, clone hRPK.998_F_8, complete sequence.
ACCESSION AC005900
VERSION AC005900.1 GI:3849824
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175066)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.998_F_8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175066)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nafé,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
DIRECT SUBMISSION
Submitted (31-OCT-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 175066)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nafé,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
DIRECT SUBMISSION
Submitted (06-NOV-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Nov 6, 1998 this sequence version replaced gi:3845400.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source
1..175066
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="hRPK.998_F_8"
/clone_lib="RPCL-17 human BAC library"
899..1372
/rpt_family="L2"
repeat_region
complement(1547..1577)
/rpt_family="CA)n"
repeat_region
1922..2133
/rpt_family="L2"
repeat_region
complement(2217..2249)
/rpt_family="CA)n"
repeat_region
4282..4579
/rpt_family="AluSc"
repeat_region
complement(5098..5129)
/rpt_family="CRAA)n"
repeat_region
complement(5130..5433)
/rpt_family="AluY"
repeat_region
complement(5922..6100)
/rpt_family="L1PA13"
6101..6182
/rpt_family="L1PI"
repeat_region
6327..6435
/rpt_family="L1PI"
repeat_region
7005..7302
/rpt_family="AluX"
7303..7613
/rpt_family="AluSp"
7615..7691
/rpt_family="A-rich"
repeat_region
complement(8284..9314)
/rpt_family="MER4-internal"
10996..11702
/rpt_family="LTR8"
repeat_region
complement(12089..12397)
/rpt_family="AluSp"
repeat_region
complement(12536..12692)
/rpt_family="AluSp/g"
repeat_region
complement(12770..13056)
/rpt_family="MER57-internal"
complement(13069..13166)
/rpt_family="HERV23"
repeat_region
complement(13194..13290)
/rpt_family="MER57-internal"
repeat_region
complement(13487..13591)
/rpt_family="MER4-internal"
repeat_region
complement(13629..13698)
/rpt_family="L1M4"
repeat_region
complement(13699..13738)
/rpt_family="TAAA)n"
repeat_region
complement(13739..14018)
/rpt_family="AluSg"
repeat_region
complement(14019..14066)
/rpt_family="L1M4"
repeat_region
complement(14067..14088)
/rpt_family="CA)n"
repeat_region
complement(14089..14209)
/rpt_family="L1M4"
repeat_region
complement(14793..15178)
/rpt_family="L1M4"
repeat_region
complement(15281..19632)
/rpt_family="L1PA2"
repeat_region
complement(19644..19752)
/rpt_family="L1P"
repeat_region
complement(19760..20454)
/rpt_family="L1ME1"
complement(20455..20615)
/rpt_family="AluY"
20616..20646
/rpt_family="TA)n"
repeat_region
complement(20647..20784)
/rpt_family="AluY"
repeat_region
complement(20785..21187)
/rpt_family="L1ME1"
repeat_region
complement(21208..21386)
/rpt_family="MIR"
21387..21688
/rpt_family="AluY"
repeat_region
complement(21689..21743)
/rpt_family="MIR"
repeat_region
complement(21821..21907)
/rpt_family="L1P"
22235..22536
/rpt_family="L1MC4"
23987..24152
/rpt_family="L2"
24233..24516

```

```

/rpt_family="Tigger2"
24517 .24634
/rpt_family="AluSg/x"
24663 .25126
/rpt_family="Tigger2"
complement(25355..25430)
/rpt_family="MERS50"
complement(25449..25627)
/rpt_family="AluSg/x"
complement(25629..26179)
/rpt_family="MERS50"
26205 26465
/rpt_family="Tigger2"
complement(26829..26951)
/rpt_family="purine-rich"
complement(26966..27083)
/rpt_family="purine-rich"
27097 .27174
/rpt_family="CCGTTn"
complement(27137..27256)
/rpt_family="G-rich"
complement(27227..27354)
/rpt_family="GGGAA"n"
complement(27329..27432)
/rpt_family="purine-rich"
complement(27450..28742)
/rpt_family="LiM4"
28882 .28932
/rpt_family="AT rich"
complement(29470..29778)
/rpt_family="AluSg1"
30002 .30029
/rpt_family="AT rich"
complement(31018..31069)
/rpt_family="LiMB3"
31070 .31140
/rpt_family="(TA)n"
complement(31148..31164)
/rpt_family="LiMB3"
complement(31165..31462)
/rpt_family="AluY"
complement(31463..31787)
/rpt_family="LiMB3"
complement(31788..32098)
/rpt_family="AluY"
complement(32099..32179)
/rpt_family="LiMB3"
complement(32274..32462)
/rpt_family="MERSA"
complement(33822..33952)
/rpt_family="LiPA5"
complement(35003..35313)
/rpt_family="AluYa5"

```

```

Query Match 3.3%; Score 20; DB 9; Length 175066;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 CTCAGTCATAAGTTTACC 77
Db 76063 CTCAGTCATAAGTTTACC 76044

RESULT 14
AC115834
LOCUS
DEFINITION Mus musculus clone RP24-15911, WORKING DRAFT SEQUENCE, 6 ordered
pieces.
AC115834
VERSION AC115834.4 GI:32567880
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 180172)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-15911
Unpublished
2 (bases 1 to 180172)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collamore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lakoque, K., Lamazares, R.,
Landers, T., Lenoczky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliker, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, N.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 180172)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collamore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafiz, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (14-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 14, 2003 this sequence version replaced gi:21362168.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24412
Center clone name: 159 L 1
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 177938 bases at least Q40

Consensus quality: 178697 bases at least Q30
 Consensus quality: 179011 bases at least Q20
 Insert size: 176000; agarose-fp
 Insert size: 179572; sum-of-contigs
 Quality coverage: 11.4 in Q20 bases; agarose-fp
 Quality coverage: 11.2 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced by the finished sequence as soon as it is available, and the accession number will be preserved.

705 704: contig of 704 bp in length
 805 804: gap of 100 bp
 1451 1450: contig of 646 bp in length
 1551 1550: gap of 100 bp
 105069 105068: contig of 103518 bp in length
 105169 105168: gap of 100 bp
 119507 119507: contig of 14339 bp in length
 119508 119507: gap of 100 bp
 119608 119608: contig of 56824 bp in length
 176232 176232: gap of 100 bp
 176332 180172: contig of 3841 bp in length.

FEATURES

source
 1. .180172
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone_lib="RP24-159L1"
 /clone_lib="RPCI-24 Male Mouse BAC"
 1. .704
 /notes="assembly_fragment"
 805. .1450
 /notes="assembly_fragment"
 1551. .105068
 /notes="assembly_fragment"
 105169. .119507
 /notes="assembly_fragment"
 119608. .176231
 /notes="assembly_fragment"
 176332. .180172
 /notes="assembly_fragment"
 clone_end:17
 vector_side:right

ORIGIN

Query Match 3.3%; Score 20; DB 2; Length 180172;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 188 TAGACCTTCTGAACCTCTCT 207
 Db 2923 TAGACCTTCTGAACCTCTCT 2942

RESULT 15

HS1042K10

LOCUS

DEFINITION
 Human DNA sequence from clone RP5-1042K10 on chromosome 22q13.1-13.2. Contains the ADSL gene for Adenylosuccinate lyase (EC 4.3.2.2, Adenylosuccinase, ASL) and 4 novel genes (one with probable rabGAP domains and Src homology domain 3). Contains ESTs, STSs, GSSs and a putative CpG island, complete sequence.

ACCESSION

AL022238

VERSION

AL022238.1

KEYWORDS

HTG: Adenylosuccinate lyase; ADSL; CpG Island.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 184180)
 Beasley, H.
 Direct Submission
 Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 On Jan 22, 1999 this sequence version replaced gi:3927920.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep -----
 Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22

RP5-1042K10 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm

VECTOR: pCRPAC2

IMPORTANT: This sequence is not the entire insert of clone RP5-1042K10 it may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone RP5-1042K10 is at 1 in this sequence. The true left end of clone RP4-591N18 is at 184081 in this sequence. The true right end of clone RP3-377F16 is at 39440 in this sequence.

FEATURES

Location/Qualifiers
 1. .184180
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="22"
 /map="q13.1-13.2"
 /clone="RP5-1042K10"
 /clone_lib="RPCI-5"
 1. .68
 /note="Alu repeat: matches 233. .276 of consensus"
 69. .380
 /note="Alu repeat: matches 1. .309 of consensus"
 381. .401
 /note="Alu repeat: matches 276. .298 of consensus"
 640. .738
 /note="L3 repeat: matches 1179. .1280 of consensus"
 complement(765. .912)
 /note="MIR repeat: matches 61. .215 of consensus"
 913. .1209
 /note="Alu repeat: matches 1. .297 of consensus"
 1225. .1402
 /note="Alu repeat: matches 140. .312 of consensus"
 complement(1419. .1459)
 /note="MIR repeat: matches 32. .71 of consensus"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 14:14:35 ; Search time 454 Seconds
(without alignments)
5698.574 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609
Sequence: 1 atgataagaatcaatgacct.....cggataccattatcaaaacta 609

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002s.*
7: Geneseqn2003as.*
8: Geneseqn2003bs.*
9: Geneseqn2003cs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	609	100.0	2370	3	AAA65395
2	30	4.9	30	3	AAA65394
3	22	3.6	22	3	AAA65393
4	19	3.1	768	7	ACN21027
5	19	3.1	855	8	ADA29053
6	18	3.0	446	4	AA522742
7	18	3.0	471	4	AA522506
8	18	3.0	496	8	ACH25483
9	18	3.0	498	8	ACH28170
10	18	3.0	621	6	ABN92747
11	18	3.0	1125	4	ABL26531
12	18	3.0	2397	7	ADA00895
13	18	3.0	3125	4	ABL26530
14	18	3.0	11320	4	ABL21028
15	17	2.8	60	6	ABN35170
16	17	2.8	192	2.8	ADG90379
17	17	2.8	264	5	AAH65860
18	17	2.8	376	4	AAH34551
19	17	2.8	439	7	ABX43841
20	17	2.8	495	6	ABK74808
21	17	2.8	687	5	AAH65859
22	17	2.8	687	7	ACA00221
23	17	2.8	700	3	AAA12925

24 17 2.8 700 8 ACD67251
25 17 2.8 700 9 ADC34675
26 17 2.8 711 9 ADC3040
27 17 2.8 715 6 ABL61139
28 17 2.8 763 6 ABL61123
29 17 2.8 776 7 ABZ18432
30 17 2.8 837 7 ACA33595
31 17 2.8 858 9 ADC3095
32 17 2.8 963 5 AAH76872
33 17 2.8 966 5 AAH24105
34 17 2.8 1127 2 AAX26368
35 17 2.8 1153 4 AAS41275
36 17 2.8 1360 6 ABN81278
37 17 2.8 1497 9 ADC93710
38 17 2.8 1701 4 AAH43488
39 17 2.8 1820 4 ABL29301
40 17 2.8 2000 6 ABZ14957
41 17 2.8 2147 3 AAZ43879
42 17 2.8 2342 4 AA160742
43 17 2.8 2351 4 AA158956
44 17 2.8 2351 6 ADB48938
45 17 2.8 2517 7 ABX06705

ALIGNMENTS

RESULT 1
AAA65395
ID AAA65395 standard; DNA; 2370 BP.
XX
AC AAA65395;
XX
DT 15-SEP-2003 (revised)
DT 09-NOV-2000 (first entry)
XX
DE Brevibacterium lactofermentum gltBD gene DNA sequence SEQ ID NO:7.
XX
KW Brevibacterium lactofermentum; ABC transporter; breeding; ATPase;
KW corynebacterium; gltBD; L-glutamic acid; GS.
XX
OS Corynebacterium glutamicum.
XX
FH Key Location/Qualifiers
CDS 1..1104
FT /*tag= a
FT /product= "ORF1"
FT 1117..1178
FT /*tag= b
FT /product= "ORF2"
FT 1759..2370
FT /*tag= c
FT /product= "ORF3"
XX
WO2000037647-A1.
XX
29-JUN-2000.
XX
16-DEC-1999; 99WO-JP007079.
XX
18-DEC-1998; 98JP-00360621.
XX
(AJIN) AJINOMOTO CO INC.
XX
Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX
WPI; 2000-452189/39.
XX
P-PSDB; AAB12591, AAB12592, AAB12593.
XX
DR
XX
PT
XX
PT
XX

ABC transporter constituent of Brevibacterium lactofermentum, its encoded gene and variants, applicable in breeding Corynebacteria particularly for production of L-glutamic acid.

PS Claim 14; Page 23-28; 34pp; Japanese.

XX The present invention describes a protein (I) which can be used to
 CC construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs
 CC can be used in breeding Corynebacteria particularly for production of L-
 CC glutamic acid. The present sequence encodes three ORFs (open reading
 CC frames) from the Brevibacterium lactofermentum gltBD gene, which is used
 CC in the exemplification of the present invention. (Updated on 15-SEP-2003
 CC to standardise OS field)

XX SQ Sequence 2370 BP; 633 A; 563 C; 515 G; 659 T; 0 U; 0 Other;
 Query Match 100.0%; Score 609; DB 3; Length 2370;
 Best Local Similarity 100.0%; Pred. NO. 8.7e-310; Indels 0; Gaps 0;
 Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGATGAAATCAATGACCTCAAGAAATCTTTGGCGTTGGATCTTATGCAAGGCTC 60
 Db 1117 ATGATGAAATCAATGACCTCAAGAAATCTTTGGCGTTGGATCTTATGCAAGGCTC 1176

Qy 61 AGTCATAAGTTTACAGAAATGACAGCACTGAGTGGAGCGTCCGGTTCAAGAAA 120
 Db 1177 AGTCATAAGTTTACAGAAATGACAGCACTGAGTGGAGCGTCCGGTTCAAGAAA 1236

Qy 121 TCAGCTTGTCACTGCTTGGCACTTGGCACTTGGCACTTGGCACTTGGCACTTGGCT 180
 Db 1237 TCAGCTTGTCACTGCTTGGCACTTGGCACTTGGCACTTGGCACTTGGCACTTGGCT 1296

Qy 181 GAGGATGTAGACCTTCTGAAATCTCTACGCTTAAGCAACGGTTATACAGGAAAATACG 240
 Db 1297 GAGGATGTAGACCTTCTGAAATCTCTACGCTTAAGCAACGGTTATACAGGAAAATACG 1356

Qy 241 GTGGGCTATTATTTCAAGATTATGCTTGAATCCCGACAGACAGTTAAATCAACCTT 300
 Db 1357 GTGGGCTATTATTTCAAGATTATGCTTGAATCCCGACAGACAGTTAAATCAACCTT 1416

Qy 301 CAGCTTGGCGTGAAGAAACACAAATGGCGTGAATTCCTCAAGTACTTTCATGCTTGGT 360
 Db 1417 CAGCTTGGCGTGAAGAAACACAAATGGCGTGAATTCCTCAAGTACTTTCATGCTTGGT 1476

Qy 361 CTTGAGTCTGTGAGAAAGCAGTTTGTGAATCTCTGTTGGCGCAACACACGAAC 420
 Db 1477 CTTGAGTCTGTGAGAAAGCAGTTTGTGAATCTCTGTTGGCGCAACACACGAAC 1536

Qy 421 GCGTGGCCCGGTACTGCTCAAAATCCCGAATAATTCCTGGCTGATGAACCAACCGGA 480
 Db 1537 GCGTGGCCCGGTACTGCTCAAAATCCCGAATAATTCCTGGCTGATGAACCAACCGGA 1596

Qy 481 GCTCTAGATTAAACAAAGTGAAGTATGATAGAGCAATTCAGAGCACTGCGCGCAAA 540
 Db 1597 GCTCTAGATTAAACAAAGTGAAGTATGATAGAGCAATTCAGAGCACTGCGCGCAAA 1656

Qy 541 GGGCCCAACCGTGTGCTGCTAGCACTCGCCCTCTTCCGAGATACAGCGATACCAT 600
 Db 1657 GGGCCCAACCGTGTGCTGCTAGCACTCGCCCTCTTCCGAGATACAGCGATACCAT 1716

Qy 601 ATCAAACTA 609
 Db 1717 ATCAAACTA 1725

RESULT 2
 AAA65394/c
 ID AAA65394 standard; DNA; 30 BP.
 XX
 AC AAA65394;
 XX
 XX 15-SEP-2003 (revised)
 DT 09-NOV-2000 (first entry)
 XX
 DE Brevibacterium lactofermentum gltBD gene PCR primer SEQ ID NO:6.
 XX
 XX Brevibacterium lactofermentum; ABC transporter; breeding; ATPase;

KW corynebacterium; gltBD; L-glutamic acid; PCR primer; ss.
 XX
 OS Corynebacterium glutamicum.
 PN WO200037647-A1.
 XX
 XX 29-JUN-2000.
 PD
 XX 16-DEC-1999; 99WO-JP007079.
 PF
 XX 18-DEC-1998; 98JP-00360621.
 PR
 XX (AJIN) AJINOMOTO CO INC.
 PA
 XX Kanno S, Kimura E, Matsui K, Nakamatsu T;
 PI
 XX WPI; 2000-452189/39.
 DR
 XX ABC transporter constituent of Brevibacterium lactofermentum, its encoded
 PT gene and variants, applicable in breeding Corynebacteria particularly for
 PT production of L-glutamic acid.
 PT
 XX Disclosure; Page 22; 34pp; Japanese.
 PS
 XX The present invention describes a protein (I) which can be used to
 CC construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs
 CC can be used in breeding Corynebacteria particularly for production of L-
 CC glutamic acid. The present sequence represents a PCR primer for the
 CC Brevibacterium lactofermentum gltBD gene, which is used in the
 CC exemplification of the present invention. (Updated on 15-SEP-2003 to
 CC standardise OS field)
 CC
 XX SQ Sequence 30 BP; 7 A; 5 C; 10 G; 8 T; 0 U; 0 Other;
 Query Match 4.9%; Score 30; DB 3; Length 30;
 Best Local Similarity 100.0%; Pred. NO. 3.5e-05;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 GCTCAACTGTCTTGGCACACTTGACAAACC 158
 Db 30 GCTCAACTGTCTTGGCACACTTGACAAACC 1

RESULT 3
 AAA65393/c
 ID AAA65393 standard; DNA; 22 BP.
 XX
 AC AAA65393;
 XX
 DT 15-SEP-2003 (revised)
 DT 09-NOV-2000 (first entry)
 XX
 DE Brevibacterium lactofermentum gltBD gene PCR primer SEQ ID NO:5.
 XX
 XX Brevibacterium lactofermentum; ABC transporter; breeding; ATPase;
 KW corynebacterium; gltBD; L-glutamic acid; PCR primer; ss.
 XX
 OS Corynebacterium glutamicum.
 PN WO200037647-A1.
 XX
 XX 29-JUN-2000.
 PD
 XX 16-DEC-1999; 99WO-JP007079.
 PF
 XX 18-DEC-1998; 98JP-00360621.
 PR
 XX (AJIN) AJINOMOTO CO INC.
 PA
 XX Kanno S, Kimura E, Matsui K, Nakamatsu T;
 PI
 XX WPI; 2000-452189/39.
 DR
 XX

PT ABC transporter constituent of Brevibacterium lactofermentum, its encoded
 PT gene and variants, applicable in breeding Corynebacteria particularly for
 PT production of L-glutamic acid.

PS Disclosure; Page 21; 34pp; Japanese.

CC The present invention describes a protein (I) which can be used to
 CC construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs
 CC can be used in breeding Corynebacteria particularly for production of L-
 CC glutamic acid. The present sequence represents a PCR primer for the
 CC Brevibacterium lactofermentum gltBD gene, which is used in the
 CC exemplification of the present invention. (Updated on 15-SEP-2003 to
 CC standardise OS field)

XX Sequence 22 BP; 5 A; 7 C; 5 G; 5 T; 0 U; 0 Other;

Query Match 3.6%; Score 22; DB 3; Length 22;

Best Local Similarity 100.0%; Pred. No. 0.56; Length 22;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 CGGACAGATCCTTCTCGAGGAT 186

Db 22 CGGACAGATCCTTCTCGAGGAT 1

RESULT 4

AC21027

ID ACA21027 standard; DNA; 768 BP.

AC ACA21027;

XX

DT 19-JUN-2003 (first entry)

XX

DE Prokaryotic essential gene #2684.

XX

KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX

XX drug design; gene.

XX

OS Acinetobacter baumannii.

XX

PN W0200277183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

XX

PR 06-SEP-2001; 2001US-00948993.

XX

PR 25-OCT-2001; 2001US-0342923P.

XX

PR 08-FEB-2002; 2002US-00072851.

XX

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

PI WPI; 2003-029926/02.

XX

DR P-PSDB; ABU17157.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway; (8)
 CC required for proliferation, or that inhibits cellular proliferation; (9)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 768 BP; 223 A; 149 C; 170 G; 226 T; 0 U; 0 Other;

Query Match

3.1%; Score 19; DB 7; Length 768;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

Qy 115 GGAAAATCGACTTTGCTCA 133

Db 115 GGAAAATCGACTTTGCTCA 133

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 8897; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the

New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
 for diagnosing a bacterial disease, as components of antibacterial
 vaccines, as targets for antibacterial drugs, or as biocontrol agents for
 plants.
 Example; SEQ ID NO 340; 328pp; English.

CC The invention relates to isolated *Acinetobacter baumannii* nucleic acids.
 CC The *A. baumannii* nucleic acids and polypeptides are useful as reagents
 CC for diagnosing a bacterial disease, as components of antibacterial
 CC vaccines, as targets for antibacterial drugs, to detect the presence of
 CC *A. baumannii* and other *Acinetobacter* species in a sample, in screening
 CC compounds for the ability to interfere with the *A. baumannii* life cycle
 CC or to inhibit *A. baumannii* infection, and as biocontrol agents for
 CC plants. The present sequence represents DNA encoding an *A. baumannii*
 CC protein.
 CC
 CC Sequence 855 BP; 239 A; 168 C; 193 G; 255 T; 0 U; 0 Other;
 CC
 CC Query Match 3.1%; Score 19; DB 8; Length 855;
 CC Best Local Similarity 100.0%; Pred. No. 20;
 CC Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 115 GGAAATCGACTTGTCTCA 133
 CC Db 199 GGAAATCGACTTGTCTCA 217
 CC
 CC RESULT 6
 CC AAS22742/C
 CC ID AAS22742 standard; cDNA; 446 BP.
 CC XX AC AAS22742;
 CC XX DT 24-OCT-2001 (first entry)
 CC XX DE Human cDNA encoding a novel human protein #308.
 CC XX KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
 CC KW immunomodulatory; cytotatic; neuroprotective; vulnery; nootropic;
 CC KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 CC KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 CC KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 CC KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 CC KW tissue regeneration; immune disorder.
 CC XX OS Homo sapiens.
 CC XX WO200155437-A2.
 CC XX PN 25-JAN-2001; 2001WO-US002623.
 CC XX PD 02-AUG-2001.
 CC XX PF 25-JAN-2001; 2000US-00491404.
 CC XX PR 25-JAN-2000; 2000US-00491404.
 CC XX PA (HYSE-) HYSEQ INC.
 CC XX PI Tang YT, Liu C, Drmanac RT;
 CC XX WPI: 2001-451939/48.
 CC XX DR P-PSDB; AAU14437.
 CC XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,
 CC XX nervous system disorders, and for regenerating bone and cartilage.
 CC XX PS Claim 1; Page 706; 894pp; English.
 CC XX
 CC The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides
 CC are used to identify compounds which bind to the polypeptides.
 CC Polynucleotides of the invention are used as probes and primers, for
 CC sequencing, for chromosome or gene mapping, in the production of
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
 CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour, in assays to determine biological activity, to raise

CC antibodies/ elicit an immune response, to determine quantitative protein
 CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC encodes a protein of the invention
 CC
 CC Sequence 446 BP; 131 A; 90 C; 91 G; 134 T; 0 U; 0 Other;
 CC
 CC Query Match 3.0%; Score 18; DB 4; Length 446;
 CC Best Local Similarity 100.0%; Pred. No. 69;
 CC Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 317 AACACAAATGGCTGAAA 334
 CC Db 157 AACACAAATGGCTGAAA 140
 CC
 CC RESULT 7
 CC AAS22506/C
 CC ID AAS22506 standard; cDNA; 471 BP.
 CC XX AC AAS22506;
 CC XX DT 24-OCT-2001 (first entry)
 CC XX DE Human cDNA encoding a novel human protein #72.
 CC XX KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
 CC KW immunomodulatory; cytotatic; neuroprotective; vulnery; nootropic;
 CC KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 CC KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 CC KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 CC KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 CC KW tissue regeneration; immune disorder.
 CC XX OS Homo sapiens.
 CC XX WO200155437-A2.
 CC XX PN 02-AUG-2001.
 CC XX PD 25-JAN-2001; 2001WO-US002623.
 CC XX PF 25-JAN-2000; 2000US-00491404.
 CC XX PR (HYSE-) HYSEQ INC.
 CC XX PI Tang YT, Liu C, Drmanac RT;
 CC XX WPI: 2001-451939/48.
 CC XX DR P-PSDB; AAU14201.
 CC XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,
 CC XX nervous system disorders, and for regenerating bone and cartilage.
 CC XX PS Claim 1; Page 273-274; 894pp; English.
 CC XX
 CC The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides
 CC are used to identify compounds which bind to the polypeptides.
 CC Polynucleotides of the invention are used as probes and primers, for

CC sequencing, for chromosome or gene mapping, in the production of
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
 CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour, in assays to determine biological activity, to raise
 CC antibodies/ elicit an immune response, to determine quantitative protein
 CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC encodes a protein of the invention

XX
 SQ Sequence 471 BP; 135 A; 96 C; 98 G; 142 T; 0 U; 0 Other;

Query Match 3.0%; Score 18; DB 4; Length 471;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 317 AACCAAAATGCGCTGAAA 334
 Db 182 AACCAAAATGCGCTGAAA 165
 |||||

RESULT 8
 ACH25483/c
 ID ACH25483 standard; cDNA; 496 BP.
 XX AC ACH25483;
 XX
 DT 13-OCT-2003 (first entry)
 XX Human adult ovary cDNA #3863.
 DE
 XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
 KW genome mapping; biodiversity; genetic disorder.
 XX Homo sapiens.
 OS
 XX US2003073623-A1.
 XX 17-APR-2003.
 XX 30-JUL-2001; 2001US-00918995.
 XX 30-JUL-2001; 2001US-00918995.
 XX (DRMA/) DRMANAC R T.
 XX (LABA/) LABAT I.
 XX (STAC/) STACHE-CRAIN B.
 XX (DICK/) DICKSON M C.
 XX (JONE/) JONES L W.
 XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
 WPI; 2003-615964/58.
 XX New polynucleotide sequences obtained from various cDNA libraries, useful
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene
 PT mapping, in the recombinant production of protein, or in generating
 PT antisense DNA or RNA.
 XX Claim 1; SEQ ID NO 12695; 44pp; English.
 XX The invention relates to an isolated polynucleotide comprising any one of
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
 CC determined by the technique of SBH (sequencing by hybridisation). Also
 CC included is a purified polypeptide comprising a sequence corresponding to

CC a reading frame of the novel polynucleotide. The nucleic acid sequences
 CC are useful in diagnostics as expressed sequence tags (EST) for
 CC identifying expressed genes or for physical mapping of the human genome,
 CC in forensics, in assessing biodiversity, or in identifying mutations
 CC responsible for genetic disorders and other traits. The nucleotide
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX
 SQ Sequence 496 BP; 138 A; 92 C; 84 G; 165 T; 0 U; 17 Other;

Query Match 3.0%; Score 18; DB 8; Length 496;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 ACAATGACAGCACTGACT 99
 Db 446 ACAATGACAGCACTGACT 429
 |||||

RESULT 9
 ACH28170/c
 ID ACH28170 standard; cDNA; 498 BP.
 XX AC ACH28170;
 XX
 DT 13-OCT-2003 (first entry)
 XX Human adult ovary cDNA #6550.
 DE
 XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
 KW genome mapping; biodiversity; genetic disorder.
 XX Homo sapiens.
 OS
 XX US2003073623-A1.
 XX 17-APR-2003.
 XX 30-JUL-2001; 2001US-00918995.
 XX 30-JUL-2001; 2001US-00918995.
 XX (DRMA/) DRMANAC R T.
 XX (LABA/) LABAT I.
 XX (STAC/) STACHE-CRAIN B.
 XX (DICK/) DICKSON M C.
 XX (JONE/) JONES L W.
 XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
 WPI; 2003-615964/58.
 XX New polynucleotide sequences obtained from various cDNA libraries, useful
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene
 PT mapping, in the recombinant production of protein, or in generating
 PT antisense DNA or RNA.
 XX Claim 1; SEQ ID NO 15382; 44pp; English.
 XX The invention relates to an isolated polynucleotide comprising any one of
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
 CC determined by the technique of SBH (sequencing by hybridisation). Also
 CC included is a purified polypeptide comprising a sequence corresponding to
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences
 CC are useful in diagnostics as expressed sequence tags (EST) for
 CC identifying expressed genes or for physical mapping of the human genome,
 CC in forensics, in assessing biodiversity, or in identifying mutations

CC responsible for genetic disorders and other traits. The nucleotide
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX Sequence 438 BP; 141 A; 96 C; 105 G; 143 T; 0 U; 13 Other;

Query Match 3.0%; Score 18; DB 8; Length 498;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 AACACAAATGGCCTGAAA 334
 DB 210 AACACAAATGGCCTGAAA 193

RESULT 10
 ABN92747
 ID ABN92747 standard; DNA; 621 BP.

XX AC ABN92747;

XX 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2210.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy; gene; ds.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-00134001.

XX 14-AUG-1997; 97US-0055779P.

XX 08-NOV-1997; 97US-0064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

XX P-PSDB; ABP40202.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermis

XX polypeptide, useful for diagnosing and treating bacterial infections.

XX Disclosure; SEQ ID NO 2210; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences can
 CC also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life cycle
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from the USPTO web site

XX Sequence 621 BP; 234 A; 94 C; 102 G; 191 T; 0 U; 0 Other;

Query Match 3.0%; Score 18; DB 6; Length 621;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 TATTATTTCAGATTAT 264
 DB 238 TATTATTTCAGATTAT 255

RESULT 11
 ABL26531
 ID ABL26531 standard; DNA; 1125 BP.

XX AC ABL26531;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31066.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Claim 1; SEQ ID NO 31066; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent,
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737,
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1125 BP; 304 A; 251 C; 291 G; 279 T; 0 U; 0 Other;

Query Match 3.0%; Score 18; DB 4; Length 1125;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 583 GAATCAGCGGATACCAT 600
 DB 826 GAATCAGCGGATACCAT 843

RESULT 12
 ADA00895/c
 ID ADA00895 standard; cDNA; 2997 BP.

XX AC ADA00895;

XX 06-NOV-2003 (first entry)

XX Mouse osteoclast related protein Gl encoding cDNA SEQ ID NO:1.

XX GI, bone resorption; osteoclast; osteopathic; gene therapy;
 KW bone related disease; osteoporosis; osteoporosis; osteoclast maturation;
 KW osteoclast related protein; gene; ss.
 XX Mus musculus.
 FH Key Location/Qualifiers
 FT CDS 45..1061
 FT /*tag= a
 FT /product= "osteoclast related protein GI"
 XX WO2003029283-A2.
 XX 10-APR-2003.
 XX 25-SEP-2002; 2002WO-EP010721.
 XX 27-SEP-2001; 2001CA-02357987.
 XX (AVET) AVENTIS PHARMA SA.
 XX Vacher J, Chalhoub N, Benachenou N;
 XX WPI; 2003-381602/36.
 XX P-PSDB; ADA00896.
 XX New GI polypeptides and genes encoding them, useful for preventing or
 PT treating bone related diseases e.g. osteoporosis or osteoporosis, and
 PT for modulating proper osteoclast maturation and bone tissue resorption.
 XX Claim 6; Fig 8; 76pp; English.
 XX The present invention describes an isolated or purified nucleic acid
 CC molecule (I) encoding a mammalian GI polypeptide which modulates bone
 CC resorption in osteoclast cells. (I) has osteopathic activity and can be
 CC used in gene therapy. GI genes and polypeptides are useful for preventing
 CC or treating bone related diseases such as osteoporosis or osteoporosis,
 CC and for proper osteoclast maturation and bone tissue resorption. Probes
 CC and antibodies raised against the GI gene product can be used in
 CC hybridisation and immunological assays to screen for and detect the
 CC presence of either a normal or mutated gene or gene product. The present
 CC sequence encodes the mouse GI osteoclast related protein, from the
 CC present invention.
 XX Sequence 2997 BP; 818 A; 642 C; 696 G; 841 T; 0 U; 0 Other;
 SQ
 Query Match 3.0%; Score 18; DB 7; Length 2997;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 77 CAGGAACCAATGACAGCAC 94
 Db 2587 CAGGAACCAATGACAGCAC 2570
 RESULT 13
 ABL26530
 ID ABL26530 standard; DNA; 3125 BP.
 AC ABL26530;
 XX
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31063.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 XX

PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX Claim 1; SEQ ID NO 31063; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 3125 BP; 966 A; 597 C; 649 G; 913 T; 0 U; 0 Other;
 SQ
 Query Match 3.0%; Score 18; DB 4; Length 3125;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 583 CAATCAGCGGATACCAT 600
 Db 1826 CAATCAGCGGATACCAT 1843
 RESULT 14
 ABL21028/c
 ID ABL21028 standard; DNA; 11920 BP.
 XX
 AC ABL21028;
 XX
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 14557.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US009231.
 PF
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.

```
PT interactions.
XX
PS Claim 1; SEQ ID NO 14557; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL161176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL161175) and the encoded proteins (ABBS57737-
CC ABB2072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11920 BP; 3521 A; 2455 C; 2509 G; 3435 T; 0 U; 0 Other;
Query Match 3.0%; Score 18; DB 4; Length 11920;
Best Local Similarity 100.0%; Pred.No. 67;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 373 GAGGAAAAGCCAGTTT 390
Db 2140 GAGGAAAAGCCAGTTT 2123
RESULT 15
ABN35170/c
ID ABN35170 standard; DNA; 60 BP.
XX
AC ABN35170;
XX
DI 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:7918.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB001503.
XX
PR 28-JUL-2000; 2000US-0221607P.
PR 02-MAY-2001; 2001US-0287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of a
PT genome, useful for detecting tissue-, pathology-, and developmental-
PT specific genes.
XX
FS Example 1; SEQ ID NO 7918; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-)
CC transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises several
CC oligonucleotides, each capable of hybridising selectively to a set of
CC messenger RNAs transcribed from a given transcription unit of the genome,
CC which encodes one or more messenger RNA splice variants. The
CC oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
```

```
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a particular
CC biological or pathological state, and so allowing the detection of tissue
CC - and pathology-specific genes such as those genes only expressed in
CC specific tissue under a specific pathological condition; to detect
CC developmental specific genes; and to detect RNA transcripts and splice
CC variants of a transcriptome of a patient suffering from a particular
CC disorder. ABN27253 to ABN59389 represent oligonucleotide sequences from
CC rats, humans and mice, which are used in the exemplification of the
CC present invention. N.B. The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 60 BP; 16 A; 15 C; 15 G; 14 T; 0 U; 0 Other;
Query Match 2.8%; Score 17; DB 6; Length 60;
Best Local Similarity 100.0%; Pred.No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 373 GAGGAAAAGCCAGTTT 389
Db 59 GAGGAAAAGCCAGTTT 43
Search completed: July 30, 2004, 15:35:53
Job time : 457 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 15:19:50 ; Search time 99 Seconds
(without alignments)
3413.789 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725
Perfect score: 609
Sequence: 1 atgataagaatcaatgacct.....cggatacattatcaaaacta 609

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	19	3.1	855	4	US-09-328-352-340
2	18	3.0	621	4	US-09-134-001C-2210
3	17	2.8	192	4	US-09-107-532A-6
4	17	2.8	700	3	US-08-718-904-109
5	17	2.8	700	4	US-09-449-249-109
6	17	2.8	711	4	US-09-107-532A-2667
7	17	2.8	858	4	US-09-107-532A-2722
8	17	2.8	1127	2	US-07-735-068-1
9	17	2.8	1497	4	US-09-107-532A-3337
10	17	2.8	2351	4	US-09-620-312B-848
11	17	2.8	2531	4	US-09-468-656A-5
12	17	2.8	2561	4	US-09-347-878-25
13	17	2.8	8195	4	US-08-961-527-94
14	16	2.6	183	4	US-09-134-001C-1676
15	16	2.6	255	3	US-08-821-994-58
16	16	2.6	285	4	US-09-312-283C-411
17	16	2.6	292	4	US-09-313-294A-7400
18	16	2.6	339	4	US-09-312-283C-410
19	16	2.6	456	4	US-09-976-594-1004
20	16	2.6	502	4	US-09-166-350-2
21	16	2.6	523	4	US-09-833-381-1256
22	16	2.6	527	3	US-09-188-930-118
23	16	2.6	527	4	US-09-312-283C-118
24	16	2.6	527	4	US-09-621-976-472
25	16	2.6	568	4	US-09-312-283C-371
26	16	2.6	585	4	US-09-252-991A-6406
27	16	2.6	653	4	US-09-833-381-1257

28	16	2.6	708	4	US-09-833-381-1258	Sequence 1258, Ap
29	16	2.6	735	4	US-09-833-381-48	Sequence 48, Appl
30	16	2.6	774	4	US-09-134-000C-520	Sequence 520, Appl
31	16	2.6	784	4	US-09-833-381-1255	Sequence 1255, Ap
32	16	2.6	804	4	US-09-252-991A-6060	Sequence 6060, Ap
33	16	2.6	846	4	US-09-543-681A-629	Sequence 629, Appl
34	16	2.6	891	4	US-09-134-001C-2153	Sequence 2153, Ap
35	16	2.6	943	4	US-09-833-381-1259	Sequence 1259, Ap
36	16	2.6	952	4	US-09-833-381-1260	Sequence 1260, Ap
37	16	2.6	1035	4	US-09-540-236-782	Sequence 782, Appl
38	16	2.6	1068	3	US-09-211-542A-11	Sequence 11, Appl
39	16	2.6	1119	1	US-07-854-596B-30	Sequence 30, Appl
40	16	2.6	1122	4	US-09-374-038-13	Sequence 13, Appl
41	16	2.6	1122	4	US-09-658-179-13	Sequence 13, Appl
42	16	2.6	1158	4	US-09-374-038-14	Sequence 14, Appl
43	16	2.6	1158	4	US-09-658-179-14	Sequence 14, Appl
44	16	2.6	1209	4	US-09-374-038-11	Sequence 11, Appl
45	16	2.6	1209	4	US-09-658-179-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-340
; Sequence 340, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 340
; LENGTH: 855
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-340

Query Match 3.1%; Score 19; DB 4; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	115	GGAAATCGACTTTGCTCA	133
DB	199	GGAAATCGACTTTGCTCA	217

RESULT 2
US-09-134-001C-2210
; Sequence 2210, Application US/09134001C
; Patent No. 6360370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2210
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2210

Query Match 3.0%; Score 18; DB 4; Length 621;

```

Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 TATTTATTTCAAGATTAT 264
Db 238 TATTTATTTCAAGATTAT 255

RESULT 3
US-09-107-532A-6/c
; Sequence 6, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Denise
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...192
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-107-532A-6
Query Match 2.8%; Score 17; DB 4; Length 192;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 GATGAACCAACCGGAGC 482
Db 116 GATGAACCAACCGGAGC 100

RESULT 4
US-08-718-904-109
; Sequence 109, Application US/08718904

```

```

; Patent No. 6037329
; GENERAL INFORMATION:
; APPLICANT: Baird, J. Andrew
; APPLICANT: Chandler, Lois Ann
; APPLICANT: Sosnowski, Barbara A.
; TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAPE
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,904
; FILING DATE: 24-SEP-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 6037329tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 760100.415C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-718-904-109
Query Match 2.8%; Score 17; DB 3; Length 700;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 AAATGCCCTGAAATTC 338
Db 314 AAATGCCCTGAAATTC 330

RESULT 5
US-09-449-249-109
; Sequence 109, Application US/09449249
; Patent No. 6503886
; GENERAL INFORMATION:
; APPLICANT: Baird, J. Andrew
; APPLICANT: Chandler, Lois Ann
; APPLICANT: Sosnowski, Barbara A.
; TITLE OF INVENTION: COMPOSITIONS CONTAINING NUCLEIC ACIDS AND LIGANDS FOR THERAI
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/449,249
; FILING DATE: 24-No. 6503886-1999
; CLASSIFICATION: <Unknown>

```

```

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,904
; FILING DATE: 24-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 650388tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 760100.415C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 109:
;
; US-09-449-249-109
;
; Query Match 2.8%; Score 17; DB 4; Length 700;
; Best Local Similarity 100.0%; Pred. No. 27;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 322 AAATGGCTGAATTC 338
; |||||
; Db 314 AAATGGCTGAATTC 330
;
; RESULT 6
; US-09-107-532A-2667
; Sequence 2667, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Weaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2667:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

QY 109 GGTTCAGGAATCGAC 125
DB 145 GGTTCAGGAATCGAC 161

RESULT 8

US-07-735-068-1
; Sequence 1, Application US/07735068
; Patent No. 5885769
; GENERAL INFORMATION:
; APPLICANT: Kumar, Chanakanti C.
; TITLE OF INVENTION: No. 5885769el Screening Systems
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 4.00B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07735.068
; FILING DATE: 19910724
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US07/655966
; FILING DATE: 14-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Dulak, No. 5885769man C.
; REGISTRATION NUMBER: 31,608
; REFERENCE/DOCKET NUMBER: ID0178K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 822 7375
; TELEFAX: 201 822 7039
; TELEX: 219165
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1127 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
US-07-735-068-1

Query Match 2.8%; Score 17; DB 2; Length 1127;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 AAATGGCGCTGAATTC 338
DB 530 AAATGGCGCTGAATTC 546

RESULT 9

US-09-107-532A-3337
; Sequence 3337, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA

ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3337:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1497 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1497
; SEQUENCE DESCRIPTION: SEQ ID NO: 3337:
US-09-107-532A-3337

Query Match 2.8%; Score 17; DB 4; Length 1497;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 AAATCGACTTTGCTCAA 134
DB 148 AAATCGACTTTGCTCAA 164

RESULT 10

US-09-620-312D-848/c
; Sequence 848, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yuning
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D

;; CURRENT FILING DATE: 2000-07-19
;; PRIOR APPLICATION NUMBER: 09/552,317
;; PRIOR FILING DATE: 2000-04-25
;; PRIOR APPLICATION NUMBER: 09/488,725
;; PRIOR FILING DATE: 2000-01-21
;; NUMBER OF SEQ ID NOS: 1105
;; SOFTWARE: pc_fg_genes Version 1.0
;; SEQ ID NO 848
;; LENGTH: 2351
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (46)..(1422)
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)..(2351)
;; OTHER INFORMATION: n = a,t,c or g
US-09-620-312D-848

Query Match 2.8%; Score 17; DB 4; Length 2351;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 AGGAAAGCCAGTTT 390
DB 1981 AGGAAAGCCAGTTT 1965

RESULT 11
US-09-468-656A-5
; Sequence 5, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-5

Query Match 2.8%; Score 17; DB 4; Length 2531;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 CTGAAATTCCTCAAGTA 345
DB 2294 CTGAAATTCCTCAAGTA 2310

RESULT 12
US-09-347-878-25
; Sequence 25, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
; FILE REFERENCE: 25885-1651
; CURRENT APPLICATION NUMBER: US/09/347,878C
; CURRENT FILING DATE: 1999-07-06
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 25
;; LENGTH: 2561
;; TYPE: DNA
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1131)..(2399)
;; OTHER INFORMATION: Escherichia coli nucleic acid encoding
;; OTHER INFORMATION: isopolyglutamate synthetase-dihydrofolate
;; OTHER INFORMATION: synthetase
;; PUBLICATION INFORMATION:
;; DATABASE ACCESSION NUMBER: M32445/GenBank
US-09-347-878-25

Query Match 2.8%; Score 17; DB 4; Length 2561;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 583 GAATCAGCGATACCAT 599
DB 1116 GAATCAGCGATACCAT 1132

RESULT 13
US-08-961-527-94
; Sequence 94, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-94

Query Match 2.8%; Score 17; DB 4; Length 8195;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 CTGAAATTCCTCAAGTA 345
DB 7895 CTGAAATTCCTCAAGTA 7911

RESULT 14
US-09-134-001C-1676
; Sequence 1676, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GIC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1676
; LENGTH: 183
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1676

Query Match 2.6%; Score 16; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 AGCTCTAGATTAAACA 495
Db 123 AGCTCTAGATTAAACA 138

RESULT 15
US-08-821-994-58/c
; Sequence 58, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
; LENGTH: 255
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 136, 187, 192
; OTHER INFORMATION: n is unknown
US-08-821-994-58

Query Match 2.6%; Score 16; DB 3; Length 255;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 547 ACCGTTGTTGTCCTA 562
Db 216 ACCGTTGTTGTCCTA 201

Search completed: July 30, 2004, 17:25:10
Job time : 100 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run. on: July 30, 2004, 15:13:30 ; Search time 2790 Seconds
(without alignments)
6518.304 Million cell up

Title: US-09-868-338-7 COPY 1117 1725

Perfect score:

Sequence: 1 atgatagaatcaatgacct.....cggataccattatcaaaacta 609

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Database :

1: em_estba:*

2: em_estbun:*

3: em_estin:*

4: em_estm:*

5: em_estru:*

6: em_estv:*

7: em_estpl:*

8: em_estro:*

9: em_esto:*

10: gb_est1:*

11: gb_est2:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estm:*

16: em_estom:*

17: em_ggs_hum:*

18: em_ggs_inv:*

19: em_ggs_liv:*

20: em_ggs_vrt:*

21: em_ggs_wtr:*

22: em_ggs_wtr:*

23: em_ggs_wam:*

24: em_ggs_pro:*

25: em_ggs_rdg:*

26: em_ggs_pbg:*

27: em_ggs_vrt:*

28: gb_ggs1:*

29: gb_ggs2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20	3.3	231	9	AV279945	AV279945 AV279945
C 2	20	3.3	244	9	AV283498	AV283498 AV283498
C 3	20	3.3	303	10	BBS22412	BBS22412 BBS22412
C 4	20	3.3	353	10	AW319130	AW319130 AW319130

5	20	3.2	10	AW210581	AW210581 um62a11.x
6	20	3.3	368	9	AI573872 uj17d06.x
7	20	3.3	379	9	AI874676 ul25a10.x
8	20	3.3	379	9	AI874676 ul25a10.x
8	20	3.3	383	9	AW012564 ul169a10.x
9	20	3.3	383	9	AW012564 ul169a10.x
9	20	3.3	392	9	AI317342 uj13e04.x
9	20	3.3	392	9	AI317342 uj13e04.x
9	20	3.3	417	9	AA397040 mx84h09.x
9	20	3.3	417	9	AA397040 mx84h09.x
9	20	3.3	427	9	AA260497 va95g10.x
9	20	3.3	427	9	AA260497 va95g10.x
9	20	3.3	440	13	BX635140 BX635140
9	20	3.3	440	13	BX635140 BX635140
9	20	3.3	445	10	B8690607 B8690607
9	20	3.3	445	10	B8690607 B8690607
9	20	3.3	459	9	AI427592 mm40h02.x
9	20	3.3	459	9	AI427592 mm40h02.x
9	20	3.3	483	10	BB747871 BB747871
9	20	3.3	483	10	BB747871 BB747871
9	20	3.3	484	9	AA250581 mw9b05.x
9	20	3.3	484	9	AA250581 mw9b05.x
9	20	3.3	478	13	B8748528 B8748528
9	20	3.3	478	13	B8748528 B8748528
9	20	3.3	478	13	BQ621602 TVEST013.
9	20	3.3	478	13	BQ621602 TVEST013.
9	20	3.3	504	9	AA374954 vb07e02.x
9	20	3.3	504	9	AA374954 vb07e02.x
9	20	3.3	504	9	AA112895 zm62f12.s
9	20	3.3	504	9	AA112895 zm62f12.s
9	20	3.3	518	9	AI182000 ud75sc08.x
9	20	3.3	518	9	AI182000 ud75sc08.x
9	20	3.3	559	9	AI166816 ui87a2.x
9	20	3.3	559	9	AI166816 ui87a2.x
9	20	3.3	600	9	AI042942 uc77e07.x
9	20	3.3	600	9	AI042942 uc77e07.x
9	20	3.3	611	9	AI035653 ud60a02.x
9	20	3.3	611	9	AI035653 ud60a02.x
9	20	3.3	748	12	BG702850 602684763
9	20	3.3	748	12	BG702850 602684763
9	20	3.3	754	9	AI326387 mm17e07.x
9	20	3.3	754	9	AI326387 mm17e07.x
9	20	3.3	774	9	AI046692 ud66h12.x
9	20	3.3	774	9	AI046692 ud66h12.x
9	20	3.3	804	9	AI255825 ui88f08.x
9	20	3.3	804	9	AI255825 ui88f08.x
9	19	3.1	249	10	B8602955 B8602955
9	19	3.1	249	10	B8602955 B8602955
9	19	3.1	306	10	B8217097 B8217097
9	19	3.1	306	10	B8217097 B8217097
9	19	3.1	311	9	AI823190 AL823190
9	19	3.1	311	9	AI823190 AL823190
9	19	3.1	419	10	Bf659278 maal3d12.
9	19	3.1	419	10	Bf659278 maal3d12.
9	19	3.1	522	29	BX124852 Danilo rer
9	19	3.1	522	29	BX124852 Danilo rer
9	19	3.1	528	13	BQ767175 EBr008.SQ
9	19	3.1	528	13	BQ767175 EBr008.SQ
9	19	3.1	544	13	BQ444036 UI-M-EW0-
9	19	3.1	544	13	BQ444036 UI-M-EW0-
9	19	3.1	544	14	CS564335 B0477D02-
9	19	3.1	544	14	CS564335 B0477D02-
9	19	3.1	552	28	BQ873918 V83F5 m7n
9	19	3.1	552	28	BQ873918 V83F5 m7n
9	19	3.1	567	28	BH750329 CH230-5AL
9	19	3.1	567	28	BH750329 CH230-5AL
9	19	3.1	580	14	CA533799 C040SG02-
9	19	3.1	580	14	CA533799 C040SG02-
9	19	3.1	585	28	B6894104 SP_ra004
9	19	3.1	585	28	B6894104 SP_ra004
9	19	3.1	596	29	CG807230 fl18079d1
9	19	3.1	596	29	CG807230 fl18079d1
9	19	3.1	600	12	B1988460 4011-41.M
9	19	3.1	600	12	B1988460 4011-41.M
9	19	3.1	600	13	BU702558 UI-M-FC0-
9	19	3.1	600	13	BU702558 UI-M-FC0-
9	19	3.1	604	14	CF257500 ph012.a0
9	19	3.1	604	14	CF257500 ph012.a0
9	19	3.1	629	14	CS563070 B0442A05-
9	19	3.1	629	14	CS563070 B0442A05-

ALIGNMENTS

RESULT 1

AV279945/C

LOCUS

DEFINITION

ACCEPTED

ACCESSION
VERSION

VERSION
KEYWORDS

KEYWORDS
SOURCE

ORGANISM

100

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 382 CCAGTTTGTGAAGTCTCTGG 401
|||||
Db 59 CCAGTTTGTGAAGTCTCTGG 40

RESULT 3
 BB522412/c
 LOCUS
 DEFINITION
 BB522412 303 bp mRNA linear EST 28-JUL-2000
 Mus musculus CDNA clone D93000724 3' similar to A7011080 Mus musculus
 mRNA for alpha-albumin protein, mRNA sequence.

ACCESSION
 BB522412
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)

REFERENCE
 AUTHORS
 Konno H., Aizawa K., Akahira S., Akiyama J., Arakawa T.,
 Carninci P., Endo T., Fukuda S., Fukunishi Y., Hara A., Hayatsu N.,
 Hirozane T., Hori F., Ishii Y., Ishikawa J., Ishikawa T., Itoh M.,
 Izawa M., Kadoto K., Kagawa I., Kai C., Kawai J., Kikuchi N.,
 Kiyosawa H., Kojima Y., Kondo S., Koya S., Kurihara C.,
 Kusekabe M., Matsuyama T., Miki R., Mizuno Y., Nakamura M., Oda H.,
 Okazaki Y., Ono T., Owa C., Saito H., Sakai C., Sato K.,
 Shibata K., Shibata Y., Shigemoto Y., Shinagawa A., Shiraki T.,
 Sogabe Y., Sugahara Y., Suzuki H., Suzuki H., Tagawa A.,
 Takahashi F., Toninaga N., Toya T., Tsunoda Y., Tataliki A.,
 Watanabe S., Yamamura T., Yamanaka I., Yano R., Yasunishi A.,
 Yokota T., Yoshida K., Yoshiki A., Yoshino M., Muramatsu M. and
 Hayashizaki Y.

TITLE
 RIKEN Mouse ESTs (Konno H., et al.)

JOURNAL
 COMMENT
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokochama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokochama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 url:http://genome.gsc.riken.go.jp/
 Carninci P., Nishiyama Y., Westover A., Itoh M., Nagaoaka S.,
 Sasaki N., Okazaki Y., Muramatsu M. and Hayashizaki Y.
 Thermostabilization and thermooxidation of thermolabile enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh M., Kikunishi T., Akiyama J., Shibata K., Izawa M., Kawai J.,
 Tomaru Y., Carninci P., Shibata Y., Ozawa Y., Muramatsu M.,
 Okazaki Y. and Hayashizaki Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)

FEATURES
 Source
 Carninci P. and Hayashizaki Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.
 Location/Qualifiers
 1..303
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="D93000724"
 /sex="mixed"
 /tissue_type="head"
 /dev_stage="15 days embryo"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, 15 days embryo
 head"

```

/Note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGGATCCAGAGACTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTTCAGTATTAATTAATTAATGCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

ORIGIN
      333      334      335      336      337      338      339      340      341      342      343      344      345      346      347      348      349      350      351      352      353      354      355      356      357      358      359      360      361      362      363      364      365      366      367      368      369      370      371      372      373      374      375      376      377      378      379      380      381      382      383      384      385      386      387      388      389      390      391      392      393      394      395      396      397      398      399      400      401      402      403      404      405      406      407      408      409      410      411      412      413      414      415      416      417      418      419      420      421      422      423      424      425      426      427      428      429      430      431      432      433      434      435      436      437      438      439      440      441      442      443      444      445      446      447      448      449      450      451      452      453      454      455      456      457      458      459      460      461      462      463      464      465      466      467      468      469      470      471      472      473      474      475      476      477      478      479      480      481      482      483      484      485      486      487      488      489      490      491      492      493      494      495      496      497      498      499      500      501      502      503      504      505      506      507      508      509      510      511      512      513      514      515      516      517      518      519      520      521      522      523      524      525      526      527      528      529      530      531      532      533      534      535      536      537      538      539      540      541      542      543      544      545      546      547      548      549      550      551      552      553      554      555      556      557      558      559      560      561      562      563      564      565      566      567      568      569      570      571      572      573      574      575      576      577      578      579      580      581      582      583      584      585      586      587      588      589      590      591      592      593      594      595      596      597      598      599      600      601      602      603      604      605      606      607      608      609      610      611      612      613      614      615      616      617      618      619      620      621      622      623      624      625      626      627      628      629      630      631      632      633      634      635      636      637      638      639      640      641      642      643      644      645      646      647      648      649      650      651      652      653      654      655      656      657      658      659      660      661      662      663      664      665      666      667      668      669      670      671      672      673      674      675      676      677      678      679      680      681      682      683      684      685      686      687      688      689      690      691      692      693      694      695      696      697      698      699      700      701      702      703      704      705      706      707      708      709      710      711      712      713      714      715      716      717      718      719      720      721      722      723      724      725      726      727      728      729      730      731      732      733      734      735      736      737      738      739      740      741      742      743      744      745      746      747      748      749      750      751      752      753      754      755      756      757      758      759      760      761      762      763      764      765      766      767      768      769      770      771      772      773      774      775      776      777      778      779      780      781      782      783      784      785      786      787      788      789      790      791      792      793      794      795      796      797      798      799      800      801      802      803      804      805      806      807      808      809      810      811      812      813      814      815      816      817      818      819      820      821      822      823      824      825      826      827      828      829      830      831      832      833      834      835      836      837      838      839      840      841      842      843      844      845      846      847      848      849      850      851      852      853      854      855      856      857      858      859      860      861      862      863      864      865      866      867      868      869      870      871      872      873      874      875      876      877      878      879      880      881      882      883      884      885      886      887      888      889      890      891      892      893      894      895      896      897      898      899      900      901      902      903      904      905      906      907      908      909      910      911      912      913      914      915      916      917      918      919      920      921      922      923      924      925      926      927      928      929      930      931      932      933      934      935      936      937      938      939      940      941      942      943      944      945      946      947      948      949      950      951      952      953      954      955      956      957      958      959      960      961      962      963      964      965      966      967      968      969      970      971      972      973      974      975      976      977      978      979      980      981      982      983      984      985      986      987      988      989      990      991      992      993      994      995      996      997      998      999      1000      1001      1002      1003      1004      1005      1006      1007      1008      1009      1010      1011      1012      1013      1014      1015      1016      1017      1018      1019      1020      1021      1022      1023      1024      1025      1026      1027      1028      1029      1030      1031      1032      1033      1034      1035      1036      1037      1038      1039      1040      1041      1042      1043      1044      1045      1046      1047      1048      1049      1050      1051      1052      1053      1054      1055      1056      1057      1058      1059      1060      1061      1062      1063      1064      1065      1066      1067      1068      1069      1070      1071      1072      1073      1074      1075      1076      1077      1078      1079      1080      1081      1082      1083      1084      1085      1086      1087      1088      1089      1090      1091      10
```

was primed with an oligo(dT) primer
[ATGCGGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DralII adaptor [GTGGGCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGTCTCTAAAGCTGG and 3' end
primer CCACCTGAGCTCGAGCACA."

ORIGIN

Query Match 3.3%; Score 20; DB 10; Length 353;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 CCAGTTTGAATCTCTGG 401
Db 163 CCAGTTTGAATCTCTGG 182

RESULT 5

AW210581 362 bp mRNA linear EST 03-DEC-1999
LOCUS um62a11.x1 Sugano mouse kidney mkia Mus musculus cDNA clone
DEFINITION IWAGS:2285948 3' similar to SM:AFAM_MOUSE 089020 AFAMIN PRECURSOR
; mRNA sequence.

ACCESSION AW210581 GI:16516521
VERSION AW210581.1
KEYWORDS EST.

ORGANISM

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 362)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.

TITLE

The WashU-NCI Mouse EST Project 1999
Unpublished (1999)

JOURNAL

COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:1008544

Seq primer: custom primer used.

FEATURES

Location/Qualifiers
1..362
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2285948"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse kidney mkia"
/note="Organ: Kidney; Vector: pME18S-FL3; Site 1: DraIII
(CACTGTGG); Site 2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGCGGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [GTGGGCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library

constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGTCTCTAAAGCTGG and 3' end
primer CCACCTGAGCTCGAGCACA."

ORIGIN

Query Match 3.3%; Score 20; DB 10; Length 362;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 CCAGTTTGAATCTCTGG 401
Db 168 CCAGTTTGAATCTCTGG 187

RESULT 6

AW210581 368 bp mRNA linear EST 29-MAR-1999
LOCUS ui17d06.x1 Sugano mouse kidney mkia Mus musculus cDNA clone
DEFINITION IWAGS:1908299 3' similar to SM:AFAM_RAT P36953 AFAMIN PRECURSOR ;
mRNA sequence.

ACCESSION AW210581 GI:4537246
VERSION AW210581.1
KEYWORDS EST.

ORGANISM

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 368)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.

TITLE

The WashU-NCI Mouse EST Project 1999
Unpublished (1999)

JOURNAL

COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:976495

Seq primer: custom primer used.

FEATURES

Location/Qualifiers
1..368
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1908299"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse kidney mkia"
/note="Organ: Kidney; Vector: pME18S-FL3; Site 1: DraIII
(CACTGTGG); Site 2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGCGGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [GTGGGCTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGTGG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCGTCTCTAAAGCTGG and 3' end
primer CCACCTGAGCTCGAGCACA."

ORIGIN

Query Match 3.3%; Score 20; DB 9; Length 368;

[illegible]

DEFINITION uj13e04.x1 Sugano mouse kidney mkia Mus musculus cDNA clone IMAGE:1907934 3' similar to SW:AFAM_RAT P36953 AFAMIN PRECURSOR ; mRNA sequence.

ACCESSION AI317342

VERSION AI317342.1 GI:4032609

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 392)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1800 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:976130

Seq primer: custom primer used.

FEATURES source Location/Qualifiers 1..392 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:1907934" /sex="female" /dev_stage="adult" /lab_host="DH10B" /clone_lib="Sugano mouse kidney mkia" /notes="Organ: kidney; Vector: pME18S-FL3; Site 1: DralII (CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTCGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCTCTTAAAGGTGG and 3' end primer CGACCTGCAGCTCGACACA."

ORIGIN Query Match 3.3%; Score 20; DB 9; Length 392; Best Local Similarity 100.0%; Pred. No. 33; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 CCAGTCTTTTGAACCTCTCTGG 401

Db 198 CCAGTCTTTTGAACCTCTCTGG 217

RESULT 10

AA397040/c

LOCUS mx84h09.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:1907934 5' DEFINITION similar to SW:AFAM_RAT P36953 AFAMIN PRECURSOR ; mRNA sequence.

ACCESSION AA397040

VERSION AA397040.1 GI:2050141

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 417)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1800 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:426649

High quality sequence stop: 393.

FEATURES source Location/Qualifiers 1..417 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:693089" /tissue_type="Liver" /lab_host="DH10B" /clone_lib="Soares mouse NML" /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5' TGTACCAATCTGAAGTGGCGCGGCAACTCTTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M.Fatima Ronaldo."

ORIGIN Query Match 3.3%; Score 20; DB 9; Length 417; Best Local Similarity 100.0%; Pred. No. 33; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 CCAGTCTTTTGAACCTCTCTGG 401

Db 346 CCAGTCTTTTGAACCTCTCTGG 327

RESULT 11

AA260497/c

LOCUS va59i0.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:747234 5' DEFINITION similar to SW:AFAM_RAT P36953 AFAMIN PRECURSOR ; mRNA sequence.

ACCESSION AA260497

VERSION AA260497.1 GI:1896999

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 427)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouse@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:456218
High quality sequence stop: 406.
Location/Qualifiers

1. 427
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:747234"
/tissue_type="Liver"
/lab_host="DH10B"
/clone_lib="Soares mouse NML"
/note="vector: pRT3D-Fac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCGAATCTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

ORIGIN
Query Match 3.3%; Score 20; DB 9; Length 427;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 382 CCAGTTTGAACCTCTCGG 401
Db 298 CCAGTTTGAACCTCTCGG 279

RESULT 12
BX635140 440 bp mRNA linear EST 12-AUG-2003
LOCUS BX635140 pBluescript Lion Mus musculus cDNA clone LIONp462D02381
DEFINITION 3', mRNA sequence.
ACCESSION BX635140
VERSION BX635140.1 GI:33615015
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 440)
Henrich, J., Hermanns, J., Kranz, H., Loebbert, R., Schluter, T.,
Schuette, D., Weindel, M., Heil, O., Ebert, L., Neubert, P., Peters, M.,
Radelof, U., Schneider, D. and Korn, B.
Mouse ArrayTAG cDNA (LION)
Unpublished (2003)
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; LIONp462D02381.
RZPDLIB;
Mouse ArrayTAG cDNA (LION)
http://www.rzpd.de/cgi-bin/products/showLib.pl.cgi?response?libNo=4
62 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heuberweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
RP: CAGGAACAGCTATGAC.

FEATURES
source Location/Qualifiers

1. 440
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="LIONp462D02381"
/lab_host="DH10B"
/clone_lib="pBluescript Lion"
ORIGIN
Query Match 3.3%; Score 20; DB 13; Length 440;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 382 CCAGTTTGAACCTCTCGG 401
Db 189 CCAGTTTGAACCTCTCGG 208
RESULT 13
BX635140 445 bp mRNA linear EST 10-OCT-2001
LOCUS BX635140 pBluescript Lion Mus musculus cDNA clone 6820448P03 3', mRNA
DEFINITION mullerian duct Mus musculus cDNA clone 6820448P03 3', mRNA
sequence.
ACCESSION BX635140
VERSION BX635140.1 GI:156017340
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 445)
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K.,
Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akashira, S.,
Tanaka, T., Tomaru, A., Toya, T., Watabiki, A., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Riken Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.
and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
source Location/Qualifiers

ORIGIN
Query Match 3.3%; Score 20; DB 13; Length 440;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 382 CCAGTTTGAACCTCTCGG 401
Db 189 CCAGTTTGAACCTCTCGG 208
RESULT 13
BX635140 445 bp mRNA linear EST 10-OCT-2001
LOCUS BX635140 pBluescript Lion Mus musculus cDNA clone 6820448P03 3', mRNA
DEFINITION mullerian duct Mus musculus cDNA clone 6820448P03 3', mRNA
sequence.
ACCESSION BX635140
VERSION BX635140.1 GI:156017340
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 445)
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K.,
Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akashira, S.,
Tanaka, T., Tomaru, A., Toya, T., Watabiki, A., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Riken Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.
and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
source Location/Qualifiers

10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Havaehizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1..463
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="F530208A04"
 /sex="male"
 /tissue_type="kidney"
 /dev_stage="adult"
 /lab_host="SOLR"
 /clone_lib="RIKEN full-length enriched, adult male kidney"
 /notes="Site_1: XhoI; Site_2: SstI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGCGCGCAACTCGAGTTTTTTTTTTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5'
 GAGAGAGAGGATCCAGAGCTCAATTAAATTAAACCCCCCCCC 3'].
 cDNA was cleaved with XhoI and SstI. "

ORIGIN

Query Match 3.3%; Score 20; DB 10; Length 463;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 382 CCAGTTTTTGAACCTCTGG 401
 |||||
Db 284 CCAGTTTTTGAACCTCTGG 265

Search completed: July 30, 2004, 17:23:28
Job time : 2797 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 07:14:14 ; Search time 3631 Seconds
(without alignments)
7269.597 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609

Sequence: 1 atgatgaaatcaatgacct.....cggataccattatcaaaacta 609

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_hgt_hum.*

31: em_hgt_inv.*

32: em_hgt_other.*

33: em_hgt_mus.*

34: em_hgt_pln.*

35: em_hgt_rod.*

36: em_hgt_mam.*

37: em_hgt_vrt.*

38: em_sy.*

39: em_hgt_hum.*

40: em_hgt_mus.*

41: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	96.4	15.8	11964	1	AE010640	AE010640 Fusobacte
C 2	91.8	15.1	341553	1	BX248355	BX248355 Corynebacte
C 3	84.6	13.9	347961	1	BX571657	BX571657 Wolinella
C 4	83.8	13.8	54310	1	AP003515	AP003515 Clostridi
C 5	83.6	13.7	301439	1	AE015943	AE015943 Clostridi
C 6	83	13.6	10575	1	AE000977	AE000977 Archaeogl
C 7	82.4	13.5	10415	1	AE007598	AE007598 Clostridi
C 8	82	13.5	12929	1	AE013394	AE013394 Methanosa
C 9	80.8	13.3	96208	1	AL596174	AL596174 Listeria
C 10	80.8	13.3	311208	6	AX417049	AX417049 Sequence
C 11	80.8	13.3	349980	6	AX413018	AX413018 Sequence
C 12	80.8	13.3	349980	6	AX417036	AX417036 Sequence
C 13	80.6	13.2	11551	1	AE006342	AE006342 Lactococcc
C 14	80.6	13.2	19841	1	SCU96166	U96166 Streptococcc
C 15	80	13.1	10738	1	AE006565	AE006565 Streptococ
C 16	79.8	13.1	10029	1	AE014953	AE014953 Streptococ
C 17	79.8	13.1	10738	1	AE010045	AE010045 Streptococ
C 18	79.8	13.1	50511	1	AE014154	AE014154 Streptococ
C 19	79.8	13.1	311650	1	AP005144	AP005144 Streptococ
C 20	79.2	13.0	225528	1	AL591984	AL591984 Listeria
C 21	79.2	13.0	244528	6	AX641673	AX641673 Sequence
C 22	78.2	12.8	94802	2	SPNEU1913	AL449935 Streptococ
C 23	77.6	12.7	10950	1	AE007663	AE007663 Clostridi
C 24	77.4	12.7	10687	1	AE008518	AE008518 Streptococ
C 25	76.2	12.5	10899	1	AF242367	AF242367 Lactococcc
C 26	75.6	12.4	21975	1	AE014236	AE014236 Streptococ
C 27	75.4	12.4	708	6	AX570190	AX570190 Sequence
C 28	75.4	12.4	10462	1	AE007458	AE007458 Streptococ
C 29	75.4	12.4	21706	6	AR218804	AR218804 Sequence
C 30	75.4	12.4	21706	6	BD003716	BD003716 Polynucle
C 31	75.4	12.4	349980	6	AX571764	AX571764 Sequence
C 32	75	12.3	12278	6	AX416895	AX416895 Sequence
C 33	74.8	12.3	12451	1	AE015017	AE015017 Streptococ
C 34	74.8	12.3	17650	1	AF397166	AF397166 Streptococ
C 35	74.8	12.3	301289	1	AE017000	AE017000 Bacillus
C 36	74.6	12.2	12834	1	U39686	U39686 Mycoplasma
C 37	74.6	12.2	13370	1	U39728	U39728 Mycoplasma
C 38	74.6	12.2	80073	6	AR300198_5	Continuation (6 of
C 39	74.6	12.2	110000	6	AR300198_0	AR300198 Sequence
C 40	74.6	12.2	250050	6	AL591978	AL591978 Listeria
C 41	74.6	12.2	349980	6	AX641667	AX641667 Sequence
C 42	74.6	12.2	349980	6	AX641668	AX641668 Sequence
C 43	74.4	12.2	12214	1	AE006299	AE006299 Lactococcc
C 44	74.2	12.2	291804	1	AE017039	AE017039 Bacillus
C 45	74	12.2	753	6	AX608453	AX608453 Sequence

ALIGNMENTS

RESULT 1
AE010640/c
LOCUS
DEFINITION
Fusobacterium nucleatum subsp. nucleatum ATCC 25586, section 182 of 197 of the complete genome.
ACCESSION
AE010640
VERSION
AE010640.1
KEYWORDS
GI:19714998
SOURCE
Fusobacterium nucleatum subsp. nucleatum ATCC 25586
ORGANISM
Fusobacteriales; Fusobacteriaceae; Bacteria; Fusobacteria; Fusobacterium.
REFERENCE
1 (bases 1 to 11964)
Kapatriai, V., Anderson, I., Ivanova, N., Resnik, G., Los, T., Lykidis, A., Bhattacharyya, A., Bartman, A., Gardner, W., Grechkin, G.,

Zhu, L., Vasieva, O., Chu, L., Kogan, Y., Chaga, O., Goltsman, E., Bernal, A., Larsen, N., D'Souza, M., Walunas, T., Pusch, G., Haselkorn, R., Fonstein, M., Kyripides, N. and Overbeek, R.
Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586
J. Bacteriol. 184 (7), 2005-2018 (2002)
2186394
J. Bacteriol. 184 (7), 2005-2018 (2002)
11889109
2 (bases 1 to 11964)
Kapral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T., Lykidis, A., Bhattacharaya, A., Bartman, A., Gardner, W., Grechkin, G., Zhu, L., Chu, L., Kogan, Y., Chaga, O., Goltsman, E., Bernal, A., Larsen, N., D'Souza, M., Walunas, T., Pusch, G., D., Haselkorn, R., Fonstein, M., Kyripides, N. and Overbeek, R.
Direct Submission
Submitted (13-FEB-2002) Integrated Genomics, 2201 W. Campbell Park Drive, Chicago, IL 60612, USA
Location/Qualifiers
1. .11964
/organism="Fusobacterium nucleatum subsp. nucleatum ATCC 25586"
/mol_type="genomic DNA"
/strain="ATCC 25586"
/sub_species="nucleatum"
/db_xref="ATCC:25586"
/db_xref="taxon:190304"
complement(98.772)
/gene="FN1352"
complement(98.772)
/gene="FN1352"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="AAL95548.1"
/db_xref="GI:19714993"
/translation="MDNREVLELVKVNKIVGDLHAKVNFQVRKGEWAIMGSSGS GKSTIMNIGMDKPSIEGVLGDITKESQSLTKIRKIGLIFQOFLPYLPTA LENVVAQYHISIDEEBALQERVLGKRAKLPQLSGGEOQRCVCIARALINSE IILADEPTGNLDEVNEKIVIEILQLKHEGSTIVVTHLEVDGVAERKILLEYGKII DIIDQK"
complement(776.1978)
/gene="FN1353"
complement(776.1978)
/gene="FN1353"
/codon_start=1
/transl_table=11
/product="ABC transporter permease protein"
/protein_id="AAL95549.1"
/db_xref="GI:19715000"
/translation="MTRKQMTIKLVVSLIRKRMIVALLAVAIGATIMSGLVIYY DIPQLGKEFFSYGANFVLPSEKIDTFDFKKNEMSTQKIVGNAPRYETTKIN CQPYLTGDMIEVKNSPFWYIEGWSNDENNVMKIEISKLNLOVGETFIIEG PRKAGVAVSQSDSAEBSKCOLNDSPYKSKLVKGIITGGAESFIFLPSILANE ILEDDTKDSTECIEADSKOLDNLANKLTADENITARPPIKVTQSDIVLGLQAL VLLNIVLILMTLSVSTWMAVAAERKEIGLAKGAVDSIRKEFLGEGSALGPI GGLGVGIGFVFAEVSLSVFGRAIEFQWLPAPITIVSMIITLACLYPVKAMEIE FALVHKGK"
complement(1988.3274)
/gene="FN1354"
complement(1988.3274)
/gene="FN1354"
/codon_start=1
/transl_table=11
/product="ABC transporter permease protein"
/protein_id="AAL95550.1"
/db_xref="GI:19715001"
/translation="MIMFMRVGRGLFRQSKMLMIAFTVALGVSLATAMNVMVGIV DKVNEKLYGANTVMHKDASILDLYLGSSEGVSNKFLSEGEVPIKQIFWGFAIV DPALBERTGVEBSVKNKIYGFNFKHLVMPTEGEVDAGIKNLKTWMEKGLKD KQVLEGTGNTKAGYKNIKLGDSINIKGTNETKLTIVRGLSGDDDEALIVLKT CQLEDEKNTINEVSALITPDNDLAKAQDPSNLTSISETYWCYAYVSSISYIQI EVLTDVAKNPRVSEEGITLNLKTEMLLILCLSSFASALGSLNLTATYSIERSOE IGLIKATGNTNRRIILLILTLVLTGIFGGIFGIAGTQIIGTKTVFSSYIEPAVI

VVPIDIALVFAVTIIGSIPAIRYLLTLKTEVLHGR"
complement(3271.4167)
/gene="FN1355"
complement(3271.4167)
/gene="FN1355"
/codon_start=1
/transl_table=11
/product="Integral membrane protein"
/protein_id="AAL95551.1"
/db_xref="GI:19715002"
/translation="WTKFVAFGEOSFGTQSLRVGGFLGLLTIFLIALSVOKVYFR LKSSIAKIFALLIYLIASDFELRGVSALARLRLKSSNLVSNLVNVPEDKSTAYIVI LTIIVACISLALLFKDSRKVIGTFKRNALLRLKARLKNKHWSLAFSLFSI LVSHTIPVATPPQIQEENMIVIFLTDVGHRLHRSYIATGNNVRFIVVKK PKGSGYGLGDACDTCGAGYFERNDEIVCKRCDVVMNKSTIGFGGCGNPVFEYELK NKIYIDKATLEKEKDRFPVGD"
4564.4722
/gene="FN1356"
4564.4722
/gene="FN1356"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAL95552.1"
/db_xref="GI:19715003"
/translation="MYDKYIKVLSIQPPFIVFYVLYNLFILWTFVHLKSPCYFF REITATLP"
4831.6006
/gene="FN1357"
4831.6006
/gene="FN1357"
/codon_start=1
/transl_table=11
/product="Transposase"
/protein_id="AAL95553.1"
/db_xref="GI:19715004"
/translation="WFLIGIDIAKLHVASCIDSTNEIVFSNFKFNDKFGFSTLLD KIKTDKNLITLGESTHYGENLINFQHFVALLINPQTSHLRANIRADKNDN LQSLNIAKSLFAKLNIFSEKNICFLKTRFRSSLIKORSKAKIOLTLLOLFFP ELQVAKSHSKAIYSLIKYFSTEEAALDDDEISLNVASKGHPKSKTELKS IAKTVIGKDSISLHVLIQLIELYLDKQIKDVTADTVDKLIVSPGISII ACAILGTNNENFSDTKLLAFAGLPKIQSGNFNASSCRMSKSKSPILRALIF TAWNIVRSEKFNKYICLRSGKSHYNSLGHVAKHVVIFTLIKKNIVYQENLE"
6403.7104
/gene="FN1358"
6403.7104
/gene="FN1358"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAL95554.1"
/db_xref="GI:19715005"
/translation="MYEDNPRLENEVLEIIMKVNFLSYNSQNTTETDPENIFEN VTTDDNKIYLPAPTDLELAKWGFPSNMDTITLNFNVVEIILENEIGLVNPPFG DLVILSKWELSELKDMKXDRKLVAEVRIEANSKILISEPKHLPMTMIETIKNCDDNLG NIKKAWLEMTKEDKSWILLIDFEGDKNYIFSKISQARNYLGNWYLDMLPYEDDFA RNSVQNHKAFYTKK"
7279.8766
/gene="FN1359"
7279.8766
/gene="FN1359"
/codon_start=1
/transl_table=11
/product="Dipeptide-binding protein"
/protein_id="AAL95555.1"
/db_xref="GI:19715006"
/translation="MKRLFFKVLVSLITLTFIACQKEENKEESIRTVSDIDSL NPYQVVSASQILINVEGLIPGVDGTVVPALASVEISEDGTTFYSIRKGVKPH NGNDMDIKDVEFSNLVNSGKLNPNTEALPENIEKILDDSHIAIYLSKPSDFIY MKEAIVPDENKHLKDIAIGTGPYKILABYOKESQKLVSKNPEYNGEKAKISTVILIS PKNSVFKULSREINFLINIDPKRIPELQKYLQSSPSNCLILSNPKPKPDDE VRKAINLADKKNVIQLAMNGKPTIYTNMSPVMSKFLWNAPEQALPEKAKQIDBEK KILPMEFTLKVNSKSKFYLDTAQSIREQDKDIGITVNLMEIETAWLSDVYTNKKYVA

SLAGSGKMPDAILRYVSTYPKNETFNENRYDVLVEAEKRTSNEAKOVQVYKEAQ
KILSEBOAAFLMDPNILIIATEKIEGFQYPIPLNFALYFKK"
8776..9693
/gene="FN1360"
8776..9693
/gene="FN1360"
/codon_start=1
/transl_table=11
/product="Dipeptide transport system permease protein
dppb"
/protein_id="AAL95556.1"
/db_xref="GI:19715007"
translation="MYIKKIFRMLLSIFSGTFSPLLELIPDPTTILGIEASAK
DLENRLQLNLSFGRYWNWLCGVQGDLSFGYKVEHVFNLILPLTISIAFI
SIFVFTMSPLSPKRNKRIKIKESILSIFSIKPSFGLIIFMLYFGLIKWA
STGVNNTWQSLILPCLVLSIPKIGWISMLHYSNLYKELREDYIKVLSYNGMKVILNF
YILKNAFLPIPTGMLLLSLITGVWIEQIPSPIGRLIVQSLVLMRDPLIQLGLIF
YTSFVLLNFIDILYSLDLPRIQVGEQ"
9693..10460
/gene="FN1361"
9693..10460
/gene="FN1361"
/codon_start=1
/transl_table=11
/product="Dipeptide transport system permease protein
dppc"
/protein_id="AAL95557.1"
/db_xref="GI:19715008"
translation="MKVLYIIILVGLIIFCIAFYQNPYKISNFLLKXPSFOHIGT
DNIGRDFESLLGTFSYIPIAPSLAGILGILGAIAGYEGVIDEPLRISIF
MSIPVILITIGIIVLNGEHSIILALFVLMPTNLNLYRGLVREKHKYIKIAKIY
GVNFRILRIIAPNIIPLVNPSTNFAGAILTEASLGILYGFQIPPYPTLGNMLNE
SQHFLLEAPFIIPLGLMILFLVYKINQISKYOEK"
10460..11161
/gene="FN1362"
10460..11161
/gene="FN1362"
/codon_start=1
/transl_table=11
/product="Dipeptide transport ATP-binding protein dppd"
/protein_id="AAL95558.1"
Query Match 15.8%; Score 96.4; DB 1; Length 11964;
Best Local Similarity 48.7%; Pred. No. 2.9e-18;
Matches 299; Conservative 0; Mismatches 306; Indels 9; Gaps 1;
Qy 5 TAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTCGGATCTTATGGCAAGGTCTCAGTC 64
Db 750 TGGAGTAAATAATGTGCTAAATATATATGAGATTTACATGCTTTAAAGAGATTAAAT 691
Qy 65 ATAAGTTTTCACAGAACATGACGACTGACTGGAGCGTCCGGTTCAGGAAATCGA 124
Db 690 TTCAAGTAAGAAAGGTAATGGGTGCAATAATGGGTTTCATCTGTTCCGGAATAATCAA 631
Qy 125 CTTTGCTCAACTGCTTGGCACACTTGACAAACCAAGTTCGGACAGATCCTTGTGAGG 184
Db 630 CTATAATGAATATATAGTTGTATGGNTAAACCAAGTATTTGGTGAAGTTATTAGATG 571
Qy 185 ATGTAGACCTTCTGAACTCTCTAGCGGTAGAACCGGTTATACAGGAAATAACGGTGG 244
Db 570 GACAGGATATTACAAAAGAAAGTCAAAATCTTTTAAACAAAATAAGAGAAATAATG 511
Qy 245 GCTATTTATTCAAGATTATCCCTTGATTCGGACAGCAGAGTAAATTCACCTTCAGC 304
Db 510 GATTAATATTCAACAATTCATTTAATCTTTATTAACTGCTCTTGAANAATGATATGG 451
Qy 305 TTGGCGTGGAAAAACAAAATGGCGCTGAAATTCCTCAAGTACTTCATGC-----TG 355
Db 450 TTGCTCAGTATTATCATAGTATACAGATGAAGAAGACATTACAGGCACCTTGAAAGAG 391
Qy 356 TTGCTCTGAGTCGTTTCAGAGAAAGCCAGTTTGTGAAGTCTCTGCTGGTGGGACACAAAC 415
Db 390 TTGGACTTTAAGATAGACAAAACATTTACCTACTCAATTATCTGCTGGTGGAGAACACAAA 331

Qy 416 GAACGTGCGTTGGCCCGGTACTGCTCAAAATCCCGGATATTTGCGTGTATGAACAA 475
Db 330 GAGTATGTATAGCAAGAGCTTTTAACTCAATAGCCCTGAAATAATATCTTGCAATGAGCAA 271
Qy 476 CCGGAGCTCTAGATTATTAACAAACAGTGTAGTGTATCATGAAGCATTTGAGACATCGCGG 535
Db 270 CAGGAACACCTTGATGAAGTAAATGAAAAATTTGTTATAGAAATATTTAAACCACTTCATA 211
Qy 536 ACAAGGCGCCACCGTGTGTGTGTAGCGACTCGCCCTCTTCGAGATCAGCGGATA 595
Db 210 AGAAGGTTTCAACAATTTGTTGTAAACATGATTTAGAAAGTTGGAGATGGGAGAGA 151
Qy 596 CCATTATCAAACTA 609
Db 150 GAAATAATATTA 137

RESULT 2
BX248355/c
LOCUS
DEFINITION
Corynebacterium diphtheriae gravis NCTC13129, complete genome;
segment 2/8.

ACCESSION
BX248355
VERSION
BX248355.1
KEYWORDS
complete genome.
SOURCE
Corynebacterium diphtheriae
ORGANISM
Corynebacterium diphtheriae

REFERENCE
1 (bases 1 to 341553)
Bacteria; Actinobacteridae; Actinobacteriaceae; Corynebacterium.

AUTHORS
Cerdeno-Tarraga, A.M., Estratiou, A., Dover, L.G., Holden, M.T.G.,
Fallen, M., Bentley, S.D., Besra, G.S., Churcher, C., James, K.D., De
Zoya, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T.,
Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M.A.,
Rabinowitz, E., Rutherford, K., Thomson, N.R., Unwin, L.,
Whitehead, S. and Barrell, B.G. Parkhill, J.

TITLE
The complete genome sequence and analysis of Corynebacterium
diphtheriae NCTC13129

JOURNAL
Nucleic Acids Res. 31 (22), 6516-6523 (2003)

REFERENCE
2 (bases 1 to 341553)

AUTHORS
Cerdeno-Tarraga, A.M.

TITLE
Submitted (03-OCT-2003) Cerdeno-Tarraga A.M., submitted on behalf
of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust
Genome Campus, Hinxton, Cambridge CB10 1SA E-mail:
amct@sanger.ac.uk

FEATURES
Location/Qualifiers

1..341553
/organism="Corynebacterium diphtheriae"
/mol_type="genomic DNA"
/strain="NCTC13129"
/db_xref="taxon:1717"
/note="biotype gravis"
complement(351..1163)
/locus_tag="DIP0382"
complement(351..1163)
/locus_tag="DIP0382"
/note="Similar to Streptomyces coelicolor hypothetical
29.8 kDa protein 2SD46.17C TR:Q9FCG6 (EMBL:AL391406) (271
aa) fasta scores: E(): 3e-26, 47.92 id in 265 aa"

gene
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"

CDS
/protein_id="CAE48886.1"
/db_xref="GI:38199255"

translation="MADHNLRLATFGHREPIGVITRGTGTFNRLRCDRWAAHPLI
QALRRTPDALDVGYSFETTVEMARLRTMDDVVTGLEIDPERVLPFRDGV
PLGLGFDLAGYOPTLVRAFNVLRQYDVQVMDAMWMTSRLOPGGFFVEGTCDELGR
STWLLNHAGPSLTLCDPDPVERPESDIASRLPKLILHRNVAGSKIHDVLLRTDEAW
DECAFWASYGRVREWARALMDMGIPMSIIRRPVRDNLTVPSIIVDPIT"
complement(1179..11640)


```
glycosyl transferase GLGA TR:Q9X9U5 (EMBL:AJ243803) (387
Query Match      15.1%; Score 91.8; DB 1; Length 341553;
Best Local Similarity 49.2%; Pred. No. 1.9e-16;
Matches 280; Conservative 0; Mismatches 277; Indels 12; Gaps 1;

QY 53 AAGTCTCAGTCATAGTTTTCACAGAAACAATGACAGCACTGACTGAGCGTCCGGTT 112
DQ 69796 AGGGAATTGATCTCACTATTTTACCTGAGAACTGTTGCAATTTGCGGAAATCTGGTT 69737
QY 113 CAGGAAATCGACTTTGCTCAACTGCTTTGCGACACTTGACAAACCAAGTCCGACAGA 172
DQ 69736 CAGGAAATCTACATTAATGACTGCTTTAAGTGATTAATCGCACCGCATGAAGGATCTA 69677
QY 173 TCCTTGTGAGGATGTAGACCTTCTGAACACTCTCTAGCGGTAAGCAACCGGTTATACAGGA 232
DQ 69676 TCTACCTAGCCAAATCAGAACATCAACCGCCACCTCGCCACGAAAGTGCCTGAGATCAGGA 69617
QY 233 AARATACGCTGGCTATTTATTTCAAGATTAATGCTTGAATCCCGACAGACAGATTAAAT 292
DQ 69616 GAAATCAGCAAGTTTCAATTTTCAAGACTTTAAACCTCATTAGTCCCTGACGCTGCCG 69557
QY 293 TCAACCTTTC-----AGCTTGGCGTGGGAAACACAAATGCGCTGAATTCCTC 340
DQ 69556 ACAACATCGGCTTCTTCAAGCTTGTGCGCCGAAACCCCAACCAACAGAAATCGACT 69497
QY 341 AAGTACTTCACTGTTTGGTCTTGAGTCTGTCGAGGAAAGCCAGTTTGTGAATCTCTG 400
DQ 69496 CAGTCTCTGAAAAAGTGGGCTGTGAGGAGTGGCAATAATATCCAAACCAATATCCG 69437
QY 401 GTGCGGAAACAAACAGAACTCGTTGGCGGGTACTGCTCAAAATCCCGGATATTC 460
DQ 69436 GAGGACACACACAGCGGTAGTATCGAGCTTCTTGTGAGTCTCCACGATCTCT 69377
QY 461 TGGCTGATGACCAACCGGAGCTCTAGATTAAACAAACAGTGAAGTATGATCAAGACAT 520
DQ 69376 TTGCGGACGAAACCGGAGCACTAGACGTTCACACCGGCAATATATCTTAAACTAC 69317
QY 521 TGAGAGCACTCGCGCAAGAGCGCCACCGTGTGTTGTTGTCAGCACTCCCGCTCTCC 580
DQ 69316 TTCAAGAAACGGTCACCAAGAGAACTTCTCTAGTTATGTCACACATGATCTGGATATTG 69257
QY 581 GAGATCAGCGGATACCATTTATCAAACTA 609
DQ 69256 CAGCATCAGTGCAGCGGTGTGTGTGCTA 69228

RESULT 3
BX571657/c
LOCUS              347961 bp  DNA  linear  BCT 08-SEP-2003
DEFINITION         Wolinella succinogenes, complete genome; segment 1/7.
ACCESSION           BX571657
VERSION             BX571657.1 GI:34482172
KEYWORDS            complete genome.
SOURCE              Wolinella succinogenes
ORGANISM            Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
                    Helicobacteraceae; Wolinella.
REFERENCE           1
AUTHORS             Schuster, S.C., Baar, C., Eppinger, M., Raddatz, G., Simon, J.M.,
                    Lanz, C., Klumek, O., Nandakumar, R., Gross, R., Rosinus, A.,
                    Keller, H., Jagtap, P., Linke, B., Meyer, F. and Lederer, H.
TITLE               Complete genome sequence and analysis of Wolinella succinogenes
JOURNAL             Unpublished
REFERENCE           2
AUTHORS             Schuster, S.C.
TITLE               Direct Submission
JOURNAL             Submitted (15-MAY-2003) Max-Planck Institut for Developmental
                    Biology, Spemannstr. 35, 72076 Tuebingen, GERMANY
FEATURES            1..347961
                    /organism="Wolinella succinogenes"
                    /mol_type="genomic DNA"

/strain="DSMZ 1740"
/db_xref="taxon:844"
1..1314
/genes="dnaA"
/locus_tag="WS0000"
1..1314
/genes="dnaA"
/locus_tag="WS0000"
/codon_start=1
/transl_table=1
/product="dnaA, probable chromosomal replication initiator
protein"
/protein_id="CAE09174.1"
/db_xref="GI:34482173"
/transl_table=1
/translation="MLGPTTLKQKDKRISTLEYEYIKQLRYDREASRTDLAVYIAPN
QRIANWKYKGERLAHPELSTGIRPKIEIKGLKQKDVSSPAAGSKGQKSTIL
NPSFTDSEVNGSNRFAYEVSQNVAKQGIAYNPLLYGGTGLGKTHLLNSIGNYV
AGKSVIYVTSQFLNDLYLHRRNTMDRFRDKYRACDYLIDDDVQFGPKQIQEEF
PHTFELHNNKQKQIVLTDKTPKQIAGLEERLKSFEWGMVSDIQPPELETINIKK
KCEFGIYLSNEIISVIATNNDNIRETEGIIKLNAYANLMOBITLQFAKNVLEQ
IKERENITLSENIIEVSKELNVKPAEIKSKGRSKNIANARIVIFLARTLTTPNMP
LAQFGMDHSSVSAMKTVKTEIIEEDANFKWVIEELKNKIKSRN"
1474..2541
/genes="DNAN"
/locus_tag="WS0001"
1474..2541
/genes="DNAN"
/locus_tag="WS0001"
/EC_number="2.7.7.7"
/codon_start=1
/transl_table=1
/product="DNA POLYMERASE III, BETA CHAIN"
/protein_id="CAE09175.1"
/db_xref="GI:34482174"
/transl_table=1
/translation="MNVIPKSSLENIITNLOPLEKKDAQSITSHVYFKAGSELTP
QATDYEIGLLAKTSVOIQEGIATVNGKILDIIVRRKKEGIEITLQSEYILHKKQEK
SFKLPMNAEEDFFSPSEELPKIEINSLHVLQSMKILPAIDNNPKFELNGALLD
IKESFNVAEDTRRLAVIKYENSPINKLIVPKKAIKVEIKLQFEEVEHYETHLL
ILKSNYLFPTKLISGKPYDERIKPETHMTLPKPKIEIKLINSLENDIKVTF
TKBELFESLSEENSEAKTQVEVETGKTKLNGINSKYLLDLQIDTSEFKIGINE
SNTFFLKSENFSTVIMPIVM"
2561..4879
/genes="gyrB"
/locus_tag="WS0002"
2561..4879
/genes="gyrB"
/locus_tag="WS0002"
/EC_number="5.99.1.3"
/codon_start=1
/transl_table=1
/product="DNA GYRASE SUBUNIT B"
/protein_id="CAE09176.1"
/db_xref="GI:34482175"
/transl_table=1
/translation="MNOQYQVADNKKVLEAVRKEPGMYIGTNNVNGLHLLIYEVV
DNSIDEAMAGHCNTISITLTKEGSIIEDNGRGIIVDMHTPTENLPAAVTLVLAGG
KFDKDSYKVGGLHGVGVVVALSKLVMTIRKNGNYTQEPSCGIPKTKGLDITGS
KNGCTTLEFPDDEIFBVTVPDREVLARRFKELAYLNRTIITHKDRDNFESYHEG
GGLRQFVHLNKKVLSQVFSFSDNESDTEVEIALAINDGDEKVSFVNNIKTPEGG
THEAGFRAGLSRAITSYIDANAREKNDITGDVREGLVAVIVSKVMEFPQEGQTK
KLGCSFVKPIQKLTYEKAKFPEENPDAKAIMQALMAARGAARAKARELTRKK
DSLMVGTLPKGLADCKPDSISELYVEGDSAGSAGKQGRDVFQAILPLKGLVAV
EKSRLDKILKSDIKMTALGCGIGEEFDINRYOKIIMTDADVDGSHIOTLLLT
FFRYLKPVTIAGVLYIAQPLVYFKKKEKELVLRDEKALSEVLIENGIEAFEGVG
TPDLLEILRASHVYELTLEKRFAMIDVVKYKWEHRDLISLMSLYIKLEFVQK
KGYNILNKRVEEKIHLVQTHGLVDLIDDSLFTPLFEEAYYVNNILRLDLSFL
GEKDILLEIEESAKKGATYQIRKGLGEMNPELQWETTTTTPENRRLKVKQITDDME
ASDVSLFMGDEVEPRRAYIQDHAKQVHLDV"
4890..5264
/locus_tag="WS0003"
4890..5264
/locus_tag="WS0003"
/notes="HYPOTHEICAL PROTEIN CUI74C"
/codon_start=1
```


ACCESSION	AP003515	CDS	3147..3509
VERSION	AP003515.1		/gene="PCP03"
KEYWORDS			/note="120 aa, no significant homology"
SOURCE			/codon_start=1
ORGANISM	Clostridium perfringens str. 13		/transl_table=11
	Clostridium perfringens str. 13		/product="hypothetical protein"
	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.		/protein_id="BAB62440.1"
REFERENCE	1	gene	/db_xref="GI:15076715"
AUTHORS	Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,H., Yamashita,A., Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.		/translation="NEKILAEKRINISFYKRNKALVTLLPPEKLEVIGITENERE KFYTEDAKIKISKEQSEAKEKTIISFKSTSTKTLNKNWLEVLGVSEDESCIIEL RKQDITLVKDNCRDILDI"
TITLE	Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater		3773..4024
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)		/gene="PCP04"
MEDLINE	21664373	CDS	3773..4024
PUBMED	11792842		/gene="PCP04"
REFERENCE	2 (Bases 1 to 54310)		/note="83 aa, similar to pir:T14710 probable transposase from Yersinia pestis (402 aa); 44% identity in 50 aa overlap
AUTHORS	Shimizu,T.		truncated"
TITLE	Direct Submission		/codon_start=1
JOURNAL	Submitted (12-APR-2001) Tohru Shimizu, Institute of Basic Medical Sciences, University of Tsukuba, Department of Microbiology; 1-1-1 Tennohda, Tsukuba, Ibaraki 305-8575, Japan (E-mail:tshimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354, Fax:81-298-53-3354)		/transl_table=11
FEATURES	Location/Qualifiers		/product="probable transposase"
source	1..54310		/protein_id="BAB62442.1"
	/organism="Clostridium perfringens str. 13"	gene	/db_xref="GI:15076717"
	/molecule="genomic DNA"		/translation="MSKDFLNYIKEQNFNDSNNVLSKDLFNDVLQEVVSAELDDM IGYEKNLNPQFVNSRNGYSKTIKSELGTITILNIRYN"
	/strain="13"	CDS	4040..4222
	/db_xref="taxon:195102"		/gene="PCP05"
	/plasmid="pCP13"		4040..4222
	/note="anaerobic pathogen for gas gangrene"		/gene="PCP05"
gene	940..1692		/note="60 aa, similar to gp:AF143819.1 transposase-like protein from Escherichia coli (402 aa); 38% identity in 60 aa overlap
CDS	940..1692		truncated"
	/gene="soj"		/codon_start=1
	/gene="soj"		/transl_table=11
	/note="250 aa, similar to pir:I40444 Spo0A activation inhibitor soj from Bacillus subtilis (253 aa); 37% identity in 250 aa overlap	gene	/product="probable transposase"
	PCP01	CDS	/protein_id="BAB62443.1"
	Para family"		/db_xref="GI:15076718"
	/codon_start=1		/translation="WISKHQININRIEDKNLNLVYASGITTRDVAGQIKALYDIISAE TVSNITNRIPELVSEW"
	/transl_table=11		4588..4746
	/product="soj protein"		/gene="PCP06"
	/protein_id="BAB62438.1"		4588..4746
	/db_xref="GI:15076713"		/gene="PCP06"
	/translation="VKKISVENIKGVAKTTSTANFACILEEKGDVLLVDLPQSNL TKLEKAYSMEDVSIADVLLDKNLDIEKVIKKTDFENIDILPSNVTIAPAEKILLDYN RSQNRKAKALEEEDKYDCLIDCPALNMTVNALCASDEVIVPIKIDKFDALDGLIE YLLDSIEIKDEFNPNLFKGFITMDSSTTVANKVIKQELKSVLGEKMFNTSITHQNIK VVESTPECPVFFSSKKARASLNYKDLSEIP"	gene	/note="52 aa, similar to pir:T43600 probable transposase from Yersinia pestis (105 aa); 56% identity in 50 aa overlap
	1751..3031	CDS	truncated"
	/gene="parB"		/codon_start=1
	1751..3031		/transl_table=11
	/gene="parB"		/product="probable transposase"
	/note="426 aa, similar to gp:AF300944.3 presumptive ParB protein from Lactococcus lactis subsplactis (242 aa); 30% identity in 266 aa overlap	gene	/protein_id="BAB62444.1"
	PCP02"	CDS	/db_xref="GI:15076719"
	/codon_start=1		/translation="WLYTTNVIIESLNISOFRKFTKTLIPFPNDVSLIKMLYLATEKVNK KWTNRNYPN"
	/transl_table=11		complement (5169..5804)
	/product="parB protein"		/gene="PCP07"
	/protein_id="BAB62439.1"		complement (5169..5804)
	/db_xref="GI:15076714"		/gene="PCP07"
	/translation="MAKKFSISEGLMNGISKTRKVEEFOAKENFKIEVINIDIKRN ENKPYEIVDISELAEDIKLNGLNHLVVRKLDNDMYELISGERRYALSGLVNEGKNE KIRPIIANDLNSATOVGRYERLNKLIPELKAVIEQGNLTIANASEFSLSSENGRV ILSLIDTWNKQEAVDLNKLIKKEHEKELEKXAYEKELEKLEKLEKKNQVSK LKSENEKLLKLDNSNLEERKEIEGQIKIEFEKKNKVEIKVQLNKLIDNSFNKLSQI TKEAKENKQKLDLSLKEKSNNEVDIKNTKENFVLVQNLKLDNSFNKLSQI NOMEKENVKAEETKAEFELEKYQKEISDLLKKL"	gene	/note="211 aa, similar to gp:AP001508.4 ABC transporter (ATP-binding protein) from Bacillus halodurans (213 aa); 49% identity in 214 aa overlap"
	3147..3509	CDS	/codon_start=1
	/gene="PCP03"		/transl_table=11
			/product="probable ABC transporter"
			/protein_id="BAB62445.1"
			/db_xref="GI:15076720"
			/translation="WNIIIEISLNKKYFDKVIKDFSLSIKKGEMIAISGRSGCKST LINMGLIEKDSGSEIIIDGVKIKINSKANKFLREKISYLFQNFALVDEETVEENL RLAKHTIKTKTKEEIRCLKVFGLGCKQNTIYELSGGQORQVARIARLMLKPSEI ILADEPTGSLDEENRDIISLLKELNESGKTIIVTHDNYVAKQADRIPT"

gene
complement (5804. .7966)
/gene="PCP08"
CDS
complement (5804. .7966)
/gene="PCP08"
/note="720 aa, similar to gp:AP001508.3 BH0280 Gene
product from Bacillus halodurans (713 aa); 23% identity in
661 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB62446.1"
/db_xref="GI:15076721"
/translation="MKKKVALILIFIIITLSPGVSVNRHTEFMKLNKLNLENF
EVTISIDRENYNTAYESI KSLDEYNGNIPFSEVDINEDVRKVINGVESKNETENH
LPIVSGRFFHKNDISYLSIDESTCQGVNDFNGKNIHEIRTIKSKLDINTEN
LFIVQENQDLKILDLSESIIVQKRVGSSQYNTETLKIILVVFIFGIFIMF
YQVLGSKYKIGIQLGHSTFVMLKRLLEVLVLMVTVLLVFFNFKTFNSLF
WKFMLEITCIVSIMIIFIVSVIIPYIKVSKITLSNIKNRPVKSIILNSIVKVL
ASILIFSNALDLSIGKGYKXVWSETKOYILPELGNDESIQSFSEMEK
BRAYLVENKOGALADPNRYEPTSMEEAKMLPEYVRETIIVNPYLKHKVYD
CNIINISEDKRILLPEKRYNPEKEILEYGYNSQEPGSSSTCSHKTAGKLNVE
QKQIKWKSQKIFSVLLDVPNEGNYVDPIVSLTESNDKLVSVYKILGYNNPFF
KIRANSEEVINGLEKYDMSVILIDPYNLDVNSTIINIKAKRVIIIFAIIVLLAV
ISIIQLNTSLYFNQNKIIVKHLGRLIYRYMNYFIMVLITWTCPLEIASLITKD
INIIVFTLIVVIELVFIIFNLSLEKNLIKVIKGY"
complement (8043. .8354)
/gene="PCP09"
CDS
complement (8043. .8354)
/gene="PCP09"
/note="103 aa, no significant homology"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB62447.1"
/db_xref="GI:15076722"
/translation="MKRNKLIKTFIASLLVSGSLGTALAYTHSSDKFEASLPGLFS
CAQSKSFYCGSQKRRATRVKVTTLVEAKDIKAKUTAHAKTYIKYKGVTEWNSYAH
L"
gene
8779. .9012
/gene="PCP10"
CDS
8779. .9012
/gene="PCP10"
/note="77 aa, similar to probable transposase from
Yersinia pestis plasmid pMT1 (402 aa); 25% identity in 158
aa overlap"
/codon_start=1
/transl_table=11
/product="probable transposase"
/protein_id="BAB62448.1"
/db_xref="GI:15076723"
Query Match
Best Local Similarity 13.8%; Score 83.8; DB 1; Length 54310;
Matches 29%; Conservative 47.7%; Pred. No. 3.7e-14;
Matches 29%; Conservative 0; Mismatches 307; Indels 18; Gaps 1;
QY 1 ATGATAGAAATCAATGACTCAGAAATCTTTGGCGTTCCGATCTATGCGAAGTCTC 60
DB 5798 ATTATAGAAATTTCAAAATTTAAATAAAAGTATTTTGATAAGGTAATATTTAAGGATTTT 5739
QY 61 AGTCATAGTTTATACCGAGAACATGACGACCTGCTGGAGCGTCCGGTTTCAGGAAAA 120
DB 5738 TCTTTGAGTATAAAAAGGTGAGATGATGCTATTAGTGAGAGAGTGGTTTGGAAG 5679
QY 121 TCGACTTTGCTCACTGTCTGGGCACCTTGGACAAACCAAGTTCGGGACAGATCCTTGTG 180
DB 5678 AGTACCCCTTTTAAATATGATAGCTTGATAGAAAAATTTGATTCAGGAGAAATAATTATT 5619
QY 181 GAGGATGTAGACCTTCTGAACTCTCTACGGGTAGCAACGGTTATACAGGAAAAATACG 240
DB 5618 GACGTGTAAAAATATAAATAATAGCAATAGCAATAGTATTTTAAAGGAAAAAG 5559
QY 241 GTGGGCTATTTATTCAGATATATGCTTGATTCCTCCGACGAGACAGTAAATTCACCTT 300

Db 5558 ATAAGTACTTATTTCAAAATTTTCCTTGGTAGTAGAAGAACTGTAGAGAGATTTTA 5499
QY 301 CAGCTTGGCGTGA--AAACCAAAATGGCCCTGAAATTCCTCAA 342
DB 5498 AGACTTGCATTAAAGCACACAAATTAATAAATACGAAAGATTTGAAGAAGAAATAAAGA 5439
QY 343 GTACTTCATGCTTGGTTCCTGAGTCGTCGAGGAAAGCCAGTTTGTGAACCTCTCTGGT 402
DB 5438 TGTTTAAATTTGTTGGATTAGGAGTGTCAAAAACTATATATATAGTATGAGTAACTG 5379
QY 403 GCGCAACCAACAACTGCTGCTGGCCCGGCTGCTCAAAAAATCCCGCAATAATCTG 462
DB 5378 GGTGAGCAACAGAGAGTAGCTATCGTAGATTAAATGTTAAAAACCAAGTCAAAATTTT 5319
QY 463 GCTGATGAACCAACCGAGCTCTAGATTAAACAACTGAGCTAGTATGATAGAACATTG 522
DB 5318 GCAGATGAACCAACAGGCTCATTAGTAGAAGAAATAGAGATTTATATATCTTTATTA 5259
QY 523 AGAGCACTCGCCGACAAAGCGCCACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 582
DB 5258 AAGAACTTAAAGAGTGGAAACAAATAATATCGTACTCATGATAATATGTTGCA 5199
QY 583 GAATCAGCGATACCATTTATC 603
DB 5198 AAACAGCAGATAGGATAATC 5178
RESULT 5
AE015943/c 301439 bp DNA linear BCT 06-FEB-2003
LOCUS Clostridium tetani E88, section 8 of 10 of the complete genome.
DEFINITION AE015943 AE015927
ACCESSION AE015943.1 GI:28204047
VERSION
KEYWORDS
SOURCE Clostridium tetani E88
ORGANISM Clostridium tetani E88
REFERENCE 1 (bases 1 to 301439)
AUTHORS Bruggemann, H., Baumer, S., Fricke, W.F., Wierzer, A., Liesegang, H.,
Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A. and
Gottschalk, G.
TITLE The genome sequence of Clostridium tetani, the causative agent of
tetanus disease
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (3), 1316-1321 (2003)
MEDLINE 22457253
PUBMED 12552129
REFERENCE 2 (bases 1 to 301439)
AUTHORS Bruggemann, H., Baumer, S., Fricke, W.F., Wierzer, A., Liesegang, H.,
Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A. and
Gottschalk, G.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2002) Goettingen Genomics Laboratory, Institute
of Microbiology and Genetics, Georg-August University,
Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
source
1. .301439
/organism="Clostridium tetani E88"
/mol_type="genomic DNA"
/strain="Massachusetts"
/sub_strain="E88"
/db_xref="taxon:212717"
92. .778
/locus_tag="CTC01981"
92. .778
/locus_tag="CTC01981"
/codon_start=1
/transl_table=11
/product="L-serine dehydratase beta subunit"
/protein_id="AA036487.1"
/db_xref="GI:28204048"
/translation="MSRNFVSFDIVGPIWIGPSSSHTAGAAFLAKLAASIAAGEIKKY


```
NEIHPMILAMNEIGYDWTILGNHEFNNGVPTLEKVASQFSGTLMCGNVYKKGDERLKG
PYNVEKAGVKVGLIGTWTNTRWDSNENLKVRYVDPVEETKVI SEIKDKVDVMA
VVRNSEEYGNKNSAVSLANACPELVAHAEHKAHDGATYNNVLVKNLAQCT
LADKIDILTKDGGYIVSDKTDVTSKIIIMWGGTKVKNYESDKLEKLSYHKVA
IADDAQIIGELKGNLVPDEBVGIPSTOIEPTMLNINEVOMYITTKANVSAAAFR
SNAMKQGIKKSASIIYFNDTLXLLVETGQKLDYMSASYNTYKPGDLTVSF
NEDVGVNYDMFSGVKVEIIDSKEPGNRIITLNRMDNSLVKDTETKLILVNNRASH
LLNPSGIFKDGSLPKLIEKDALDGTPIRLIGRYKEVKKGVITPQMNKNKISGN
WDSAKRKEAVKLINEGXIKVPOSKDARTPNVASITKOLENAKNKAVKPVDKKPTTK
PDSKSDKDKVIRPADKXIVYVYKSGDCLYLIGQKNVSYKTIKANNIKWNLIIVFG
ORLIIPAK"

Query Match      13.7%; Score 83.6; DB 1; Length 301439;
Best Local Similarity 48.3%; Pred. No. 6.6e-14;
Matches 282; Conservative 0; Mismatches 284; Indels 18; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTGGCGTTCGGATCTTATGGCAAGGTCTC 60
DB 125348 ATTGTAAGATGGAGATATTAACATAAAATTTGGAGATAAATATCTAAATAATTTT 125289
QY 61 AGTCATAGTTTATACCAAGAACATGACAGCACTGACTGAGCGTCCGGTTCAGGAAA 120
DB 125288 TCAGTAGATATACAGATGGAGAACTCTAGCTGTAAGCAAGTGGAAAGTGGAAA 125229
QY 121 TCGACTTTGCTCAACTGCTTGGACACTTGCACAAACCAAGTTCGGACAGATCCTTGC 180
DB 125228 TCTACTATATTAATATATAGGCGTTTGTAGAGGGTTTGTAGTCTGGAAGCTTATTTA 125169
QY 181 GAGGATGATAGCTTCTGAAACTCTCTAGCGTGAAGCAACGGTTATACAGGAAAAATACG 240
DB 125168 GATGATGATGAATATATAAATAAATTCATCGAATCAATAGATATTAAGAGAAA 125109
QY 241 GTGGGCTATTTATTTCAAGATTATCCCTTGATTCCTCGACAGCAGATTAATTCACCTT 300
DB 125108 ATAGGATATTTATTTCAAAACTTTGCAATAGTAGATGAAGAAACAGTATATATAATCTA 125049
QY 301 CAGCTTGGGCTGG-----AAAAACAACAATGGCCTGAAATTCCTCAA 342
DB 125048 CATTTAGCACTTAATATATGTTAAGAAAATAAAGAAGAAAGATGATGATGATAAAAAT 124989
QY 343 GTACTTTCATGCTGTGTGGTCTGAGTCGTTGAGGAAAAAGCCAGTTTGTGAACCTCTCTGGT 402
DB 124988 GTATTAAGCAAAATGAATTTAGAGGATGATGAAGAGAAAATCTTTGAGTTAAGTGGC 124929
QY 403 GCGGAACACACAGCACTCGTGTGCGCGGCTACTGCTCAAAATCCCGAATATATTCG 462
DB 124928 GGAGAGCAACAAGAGTTTCAATAGCAAGATTTATTTAAACCAAGCAAGATATTTA 124869
QY 463 GCTGTGTAACCAACCGAGCTCTAGATTTAAACAAACAGCTAGCTAGTATAGAACGATG 522
DB 124868 GCTGTGTAACCAACCGGATCATTAGATGCTAAAAATAGAGACTTAGTATTACTATTTA 124809
QY 523 AGAGCACTCGCGCAACAAAGCCGACCGTGTGTGTGCTACGCA 566
DB 124808 AATAAACTCAACAAAGAGGAAAAACAGTAATAGTAGTAACCTCA 124765

RESULT 6
AE000977 10575 bp DNA linear BCT 17-NAR-2003
LOCUS Archaeoglobus fulgidus DSM 4304 section 130 of 172 of the complete
DEFINITION genome.
ACCESSION AE000977 AE000782
VERSION AE000977.1 GI:2689300
KEYWORDS Archaeoglobus fulgidus DSM 4304
SOURCE Archaeoglobus fulgidus DSM 4304
ORGANISM Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
          Archaeoglobaceae; Archaeoglobus.
REFERENCE 1 (bases 1 to 10575)
AUTHORS Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,
        Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,
        Richardson, D.L., Keriavage, A.R., Graham, M.D., Adams, M.D., Loftus, B.,
        Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,
        Kirkness, E.F., Dougherty, B.A., McKenney, K., Badger, J.H., Glodek, A.,
        Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,
        Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,
        Utterback, T., Cotton, M.D., Spriggs, T., Artlich, P., Kaine, B.P.,
        Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,
        Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,
        Woese, C.R. and Venter, J.C.
        The complete genome sequence of the hyperthermophilic,
        sulphate-reducing archaeon Archaeoglobus fulgidus
        Nature 390 (6658), 364-370 (1997)
        98049343
        9389475
        2 (bases 1 to 10575)
        Klenk, H.P., Clayton, R.A., Tomb, J.-P., White, O., Nelson, K.E.,
        Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,
        Richardson, D.L., Keriavage, A.R., Graham, M.D., Adams, M.D., Loftus, B.,
        Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,
        Kirkness, E.F., Dougherty, B.A., McKenney, K., Badger, J.H., Glodek, A.,
        Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A.,
        Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,
        Utterback, T., Cotton, M.D., Spriggs, T., Artlich, P., Kaine, B.P.,
        Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,
        Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,
        Woese, C.R. and Venter, J.C.
        Direct Submission
        Submitted (15-DEC-1997) The Institute for Genomic Research, 9712
        Medical Center Dr, Rockville, MD 20850, USA
        In order to show the genes in ascending order on the genome, the
        origin of this version has been moved by TIGR to position 2093570
        of the original version and the opposite strand is shown from the
        original version.
        On Dec 16, 1997 this sequence version replaced gi:2648725.
        Location/Qualifiers
        1..10575
        /organism="Archaeoglobus fulgidus DSM 4304"
        /mol_type="genomic DNA"
        /db_xref="taxon:224325"
        complement(318..1667)
        /gene="AF1813"
        complement(318..1667)
        /gene="AF1813"
        /note="similar to GP-2225877 percent identity: 47.43;
        identified by sequence similarity; putative"
        /codon_start=1
        /transl_table=11
        /product="TBP-interacting protein TIP49"
        /protein_id="AAB89434.1"
        /db_xref="GI:2648730"
        /translation="MAGEIRBITOTFERISAHSHIRGLDENLRADKVDGLVGOKR
        AREAGVIRLLKSGRMAGRGLMAGPOTGKTATNAVALSKELGKIDIPVOVSASEFY
        SAEMKTEALIQMRKAIGVIRRETITVLEGEVGLDINWVNPYNPTCKIPESATLT
        LATKDEKRTFSVSGRLAMQFTQGIIEVGDIVIDKETGIRGKLGSEKAKKIKDGLDD
        EVVPVSGKVEKEFTYVVTLDLDEANARTSIFSPSPESRIDNEVRBADEQV
        KLVBERGRLVPGVLFIDETLMDIELFAFMNAMESMAPITIIANRGRSFRIGT
        DIVAPHGILDLDLRLITTPYREBIKTIETIRAAESGIMLSNEAKELTDIGEK
        TSLRYAQLLAIPAYEFKRNKSGKVELEVERAASIFADVSSQSSAYLKKWEEKMLCM"
        1751..2107
        /gene="AF1814"
        1751..2107
        /gene="AF1814"
        /note="hypothetical protein; identified by GeneMark;
        putative"
        /codon_start=1
        /transl_table=11
        /product="A. fulgidus predicted coding region AF1814"
        /protein_id="AAB89439.1"
        /db_xref="GI:2648735"
        /translation="MKLRGHLILCLOFRGEGYSGFVENVKIMERLESGETVEVVA
        GVDDICVACPYNRGYCGKDETSEERVREMDKAAVRLGLVSEKTTTREVAAERIEKVLNE
        WRVFCDDCKWEKVCYG"
        2144..3418
        /gene="AF1815"
```

CDS	2144..3418 /gene="AF1815" /note="similar to GB:L77117 SP:Q58131 PID:1591438 percent identity: 36.19; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="acetylornithine aminotransferase (argD-2)" /protein_id="AAB89433.1" /db_xref="GI:2648729" /translation="MGFTCRGKABII EAFSKHVSYPYKAKFFSMVGIDFVPRREGVW YWLDGRKLMCHNGGVNGLGHHPETVKTIVLALDELIGNHLLISEQRAIAEKL ALMEPGDLSRVVFGGGEAIDFAIKLARGHTGKKLIYAKGVGHGTGEARAGAEK YKPEPIAPGVEVPEFGDAEAVEKAVDDPTAAVLFPATLGNPDPDPFYRVRE ICDEKGCIMINDEVTQGLGRTGCKWGHYKHYKVEDIVTAKGLSGGVPISATCFREG LDDFAENPFRHVSFTFGAELGCVVAEKVLEITSRESFLENVRKTGEALSEILGKLKD EKYVDYIRQKGLGIGIKWVEEGWGPLLSISYHSGILAVYANNDTSMVQFLPLPIVG EKVYDINREKLAGAMEIASORREVMVEFIRKML" 3415..4317 /gene="AF1816" 3415..4317 /gene="AF1816" /note="hypothetical protein; identified by GeneMark; putative" /codon_start=1 /transl_table=11 /product="A. fulgidus predicted coding region AF1816" /protein_id="AAB89438.1" /db_xref="GI:2648734" /translation="MKVDVEVLRNFERLNPAKENTKII GYGEISTVFELPELPQIA LKRNPKSVQVRYIEVYFXYIELLEGAGLRVPDSDATFVRRDQKTVLYIAQWV GNGKVIQAKADEAEVEFGRVLAENKKVWDFNAERREGLSIGDQGISNWAIMEE IYPTDTPMIRKNGIHQLOTEVFLQACPPGIRILLKKI FLQIDLRYPDFRVVJDM IANLYKGEKREDLIPIALNANEPKNGYDFEAITAKEVEKYKEDAFINSLYLNLRK VHRFILKALFGREYIILPGKRR" 4457..4966 /gene="AF1817" 4457..4966 /gene="AF1817" /note="similar to PID:1742672 PID:1742680 percent identity: 24.55; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="transcriptional regulatory protein, TetR family" /protein_id="AAB89432.1" /db_xref="GI:2648728" /translation="MTDQSVRIIEALRLYMKPPHEVSEIEEAREAKVSKSLIFVHF ESKQKLEEAIVMAFKMBEFPNRSVEEVDYIGFTIARERREFIFMMALSQVRIE ELRMFGAELEKVASLFEFGCRHPRETAIALMAMLDGLSIYSLYFDLGLKLEYRIAME FVSRVRRA" 4963..6399 /gene="AF1818" 4963..6399 /gene="AF1818" /note="hypothetical protein; identified by GeneMark; putative" /codon_start=1 /transl_table=11 /product="A. fulgidus predicted coding region AF1818" /protein_id="AAB89437.1" /db_xref="GI:2648733" /translation="MRFVLALLAFALAAQATAQPVITLNHHTIPENPSIGYFIISID ISNAGNARGLSYVENEEGFYLIINPDSEKTSVYLDGLDSAGTSQAQIKAYAEK SGLYALKASLYRNESGEFIEGVFIADVLPFIYAGSTIEIRPDESREAEVLMRL LGDVSFAVILNSDFTITSDALIESWRETGKVVGLKADSDAVAGVYEAKIIVDY RSFGKGVAEVPLVIVSEPTVLEVTFFPEKVADDFSLKVAITAKNSKLSGLT ASLEABRFDEKKYVGTVEAGKAVINFSLSKSPRESGEVDVVLVLTESFAKNRSVP LVTSYGRISIDLAGVYTPQVVEGDSFKLSIQVENSSEQDAKAVAVNLIIPESLEG BRSYFYGSLSGSSASATATEMRALKAGSQKIVAEIITMDKITEFKHKAEREFEVYVFA DRSGVLAGLALLIGVVAALAWLMRREKA" 6400..7095 /gene="AF1819" /note="similar to GB:L77117 SP:Q58206 PID:1591493 percent identity: 51.13; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="ABC transporter, Atp-binding protein" /protein_id="AAB89431.1" /db_xref="GI:2648727" /translation="MLRLDVMKVMQGVKVEVSALRGINLEIERGEFMMVVLGPSCGCK TWMNLIGIDRPTGRVIFDGDITNNEDSLTMHRRNNVGFIFQFNLIPLTFARE NVIEADLVESPRDVEDLVMGLADRAHEHFAELSGGQQRVAARALVKNPPIILA DEPTGSLDEFTCKAVLKNVREINRKEGITFVLVTHNSAIAADRVVYLRDQKVERVE RNLHPADPEIQW" 7031..9406 /gene="AF1820" 7031..9406 /gene="AF1820" /note="hypothetical protein; identified by GeneMark; putative" /codon_start=1 /transl_table=11 /product="A. fulgidus predicted coding region AF1820" /protein_id="AAB89436.1" /db_xref="GI:2648732" /translation="MERLUGRGISIRLTPMKSSGSIILRKLRLDIKQKGLFAATTF ILFGVTIFCAPYLSYLNINDTYESFSSNFEDISKAVEITEEBEIRVKKLDGVLE VPRISVGVKVDGKVEALVISLPEKETNRLYVAGEGFVAKVFPADYVIEVN TLELYSFAQVYKVEALVYSPFELIVEEGEETTPSPGYAIVYIPEKDKMLKAGL KENELKIRVAKERNEEILRVLITLGDVKVEFHTGENQSKLKEDEGRGLAVL PFAFIVSIFMTYALLSRIPLQGNANVRALGFTSNEIMLHYQLPFLMNGFFAAS AGLVGFPASQLLTQYITFNLPIYVSKHDEVSLSMAGTLTPTISGFLVFAAS RVDIVKALGYAEVAASVFIARIDALFRIWMRLIFRLALNIFRSKRRTAISIFI VACTSLIINSMVVDSPYVMQVQFYAYDIKVSLEGYDQKVEKVRKMDGVLFVA BEVEMPIYVEKGEAEVDVEMTPLLVSATSDYFSRISGVGEGNIRIVVDADEGR VEGEKVSVTFEGKLEAEVDVEMTPLLVSATSDYFSRISGVGEGNIRIVVDADEGR IAEIAEKIROMDGVKSVTLEAOFSIEELMGFFYAFIAPSLFCVSLGPRAFVNTTS ISVIERSELATLMGYTSREIILSILENLFVALLGLVPAIPATIAITAFPPFSFE SELYMPMVIPRTFAIVLAVLAILLALPLSARRVEMDIARKVTKIEVS" Query Match 13.6%; Score 83; DB 1; Length 10575; Best Local Similarity 49.1%; Pred. No. 4.3e-14; Matches 249; Conservative 0; Mismatches 255; Indels 3; Gaps 1; QY 55 GGTCTCAGTCATAAGTTTTTACCAGAACATGACAGCACTGACTGGAGCGTCCGGTTCA 114 Db 6466 GGGATAAATCTTGAATAAGAGCGGAGAGTTCATGTCGTCTTGGGCCCTTCAGCGTCG 6525 QY 115 GGAATAATCGACTTTCCTCAACTGCTTGGCACACTTGACAAACCAAGTTCGGACAGATC 174 Db 6526 GGTAAAGACGACGATGCTGACATAATTGGGGGAATTGACAGGCCAACAAAGGGGAGGTG 6585 QY 175 CTTGTGAGGAGTAGACCTTCTGAACTCTCTACGGCTAAGCAACGGTATTACAGGAAA 234 Db 6586 ATTTTGTGTTAAGGATATCACGAATTAACATGAAGACAGGCTCACCATGCACAGGAGA 6645 QY 235 AATACGGTGGGCTATTTTATTTCAAGATTATGCTTGTATTCCTCCGACAGCAGATTAATTC 294 Db 6646 AATAACGTGTGCTTCATTTTTCAGTTCTTCAACCTCATTCACACATTAACATGCAAGGAA 6705 QY 295 AACCTTCAGTTCGGGTGGAAAAACAAATTTGGCC---TGAATTCCTCAAGTACTTCAT 351 Db 6706 AACGTGGAGATTGCAGCAGATTTTGGTTCAGAGCCCGAGGATGTGGATGAGTCTGAAG 6765 QY 352 GCTGTGTGCTTGTAGTGGTTCGAGGAAGACCACTTTTGGTGAACCTCTCTGGTGGGGAACAA 411 Db 6766 ATGTCGCGACTTGTGACAGCGCGGAGCAGCTTCCCGGAGAGCTGAGCGGTGGAGAGCAG 6825 QY 412 CAACGAACTGCGTTGGCCCGCGGTACTGCTCAAAAATCCCCGAAATAATTCGGCTGATCAA 471 Db 6826 CAGAGAGTTGCCATAGCAAGGGCTCTGTTTAAGAACCGCGCAATAATTCCTCGCTGACGAG 6885
-----	---

Qy 472 CCACCGGAGCTCTAGATTATCAAAACAGTACGCTAGTACATAGAACGATTCAGACACTC 531
|||||
Db 6886 CCAACTGGCAGTCTGGACTTCGAACTGGAAAGGCTGCTGAGGTCATGAGACAGATA 6945
|||||

Qy 532 GCGACAAAGCGCCACCGTGTGTT 558
|||||
Db 6946 AACAGGAAGGAGGAATAACCTTTGTT 6972
|||||

RESULT 7
AE007598
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLES
JOURNAL
FEATURES
source

10415 bp DNA linear BCT 27-JUL-2001
Clostridium acetobutylicum ATCC824 section 86 of 356 of the
complete genome.
AE007598 AE001437
AE007598.1 GI:15023711
Clostridium acetobutylicum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
1 (bases 1 to 10415)
Nolling, J., Breton, G., Omelchenko, M.V., Markarova, K.S., Zeng, Q.,
Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I.,
Tatousov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P.,
Daly, M.J., Bennett, G.N., Koonin, E.V. and Smith, D.R.
Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
21359325
11466286
2 (bases 1 to 10415)
Childress, D., Zeng, Q. and Smith, D.R.
Direct Submission
Submitted (24-JUL-2001) GTC Sequencing Center Production,
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA
Location/Qualifiers
1. 10415
/organism="Clostridium acetobutylicum"
/mol_type="genomic DNA"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
122..724
/gene="CAC0821"
122..724
/gene="CAC0821"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator (Tetr/AcrR family)"
/protein_id="AAK78797.1"
/db_xref="GI:15023712"
/db_xref="GI:15023713"
/translation="WQYLEKREKRIIDALKFELRLCYTNASVTTIVQAGTSGVNF
KYFGKEDLYEKIDIPYKLYLHQFSEVNEQAQAFELMKVIEIPEBND
ELSLNLSGSKYENCRCQIFIDFTNTVTSFEVLYMNRKTIKDNFIELAHNV
DQAIILRRKKQKQVRKLVNLIDIFYGNIIDIKIDIOEI"
744..1448
/gene="CAC0822"
744..1448
/gene="CAC0822"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAK78798.1"
/db_xref="GI:15023713"
/translation="WESPIELRDVREKIYMGSEVITIKAVDGVFSIDEGFVIVIGASG
AGKSTNLNLGMDQASGSIISVQNEI.SNNYKMLTKYREDIDIGFVQFVNLQVNL
ALENLELVEIKENPMDPVKLVESGLKDRMNPSPQSGEQORVSIARALANPKL
LICDETALDINIGKALKLADTSKQYNMIVLITHNSAIFADIKLIHVKNRVE
SVKKNQTPVFIIEVIEW"
1450..4800
/gene="CAC0823"

CDS
1450..4800
/gene="CAC0823"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/protein_id="AAK78799.1"
/db_xref="GI:15023714"
/translation="MKKTFKFLPKSITSLRFLSILIIIAVGSFYVGVGRASSPDM
KMSADKYFNDNFMDKLIISTGLTKDVKRKLKSVKAAKGSKSDMAVINGKDTI
VLNBSLPKQGVNVKQVVRGLRPSKNTLAEVERFLKFNKIGDVLKLSGNDINI
EORLNCSPKIVGTVTSPLVISAORQISSVGNVGVKGVVIMPEAFKNDVYTSIYVKI
NEPKSDSLAONAKYKASIVNKKLKTIVKGEARERKQIVKSDSKIDKSDKLSA
SKERAKQFIDAKKLDVAVYKNDGVQLQKQNDKFTQITSNENIIONKSEIQOS
ETIINSKLADAKSKSTLEGAEQIVLQQLNDKQHAAGEGISAAUGKAVSDAKNL
KNPSDITCKEINYNLNDVYTDIQGKEFDMDYESLSSSLEKISSYFDIKGAKAF
DSAQSEISIKQTTIQDNKALISGEVQLQNALQINAGDKIAKGETELSKNKGLE
KLNAARKIDDSKEINENTKLTETNANNKFNNAEGRKRNDRKLNKREKPTTV
LGRNENLVNYSQSDRIDNIGKVFPLIFLVAALVSLTMTVMQENETGTEIGTFA
LQSPKIVSHYLVTSFTASIIIGSLGIIIGFVEPPLIVNAGSLVAVPHYSFDSV
TSIEAVITAVFTVASAIVLDELKSPASINRPPKPKAKILLERMTFIWKRFSE
TKKTARNIRPKQFPLMTVIGIACTGLMTGEGIKSGVIGAVEKQFSSVYVDQOT
TINKIDDSERKEIKDQVSDSNISVLFYISKQATVRKEDNNEDINIVPEKKQAL
NEYELFNDSELKLDSDGVIVITKIAKLMNKKGVFNITVDNRILIRVIFDITKHY
IQHYIMSPKYKVTGENAQFNAPYGLKSKSSAENTTSKISGINKAVGFPKN
TRDFNKSLNSVVLVIVSAGLAFVIVNLININERRELATIKVGFYDKE
LAMVYIRENIILTIIGSLVIGVGMILLDFVITTTETNMVFLRTVEPLVILISIFLT
IAFSVIVNIAMVRRFYKIDMIESLSKSAE"
4906..5226
/gene="CAC0824"
4906..5226
/gene="CAC0824"
/codon_start=1
/transl_table=11
/product="Thioredoxin"
/protein_id="AAK78800.1"
/db_xref="GI:15023715"
/translation="WESLQSIKEIRKYIENNELVLMYFRSDNSGICVYALDPLEATSK
KYSKISSVKNVTGNLPSVPADFKITVPEIIFMNGSIVIEHFKYINFRIEIEVLRY
YKCI"
complement (5283..6416)
/gene="CAC0825"
complement (5283..6416)
/gene="CAC0825"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78801.1"
/db_xref="GI:15023716"
/translation="MFRFLFSKVKILFLALTLGLTYLTNLNTVATNNTTKTKKIT
SLQIVANNKVGLDLGNLSPFNETDMGNPKTKSMITKIKQAGFNTPRIPITWNSHI
GPADYITDNLDRVOEVNVAVDNHYVILNHEDSWLPTIYANKDAATAEKTKL
WTOI.SNPKYNDLSLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRTG
AKNSRSPMIPYAAASDKISINGKILPNNDKILSLHAYIPYEMVNGTSSWGS
AKDKELSLTALYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA
PIWMDGVYSGGADSYGIFNRNTLRWECPRIVHALTGLKVKK"
complement (6748..7860)
/gene="CAC0826"
complement (6748..7860)
/gene="CAC0826"
/codon_start=1
/transl_table=11
/product="Endoglucanase family 5"
/protein_id="AAK78802.1"
/db_xref="GI:15023717"
/translation="MFKNLFSKAKISLLTFAFLGTSLSGAYSACAATVPHDMTSQOI
VNDKVKWNLGNTLDSDEFTGNPKTKTKMIDIKIKEAGFNTPRIPVSNSHIGAGP
SYTIDQALNRVQEVNVAVDNHYVILNHEDSWLPTIYANKDAATAEKTKL
ANRFDVSHLIPETNPELIGTPEEWYGGTSEARDVNVNLAANAIRTG
SRFIMVYTAASTAAMDVLIPNDKRVIVSLNHYKPYSPANDPKGTSHWGGEADK
DALDQGLNALYKFKGKGVVIGCEPASVKNKNENORASLAKYVVSAAKRSIA
PIWMDGVYSGGADSYGIFNRNTLRWECPRIVHALTGLKVKK"
complement (8144..9007)
/gene="CAC0827"

CDS	complement (8144..9007) /gene="CAC0827" /codon_start=1 /transl_table=11 /product="Fructose-bisphosphate aldolase" /protein_id="AAK78803.1" /db_xref="GI:15033718" /translation="MLVSAKEMVDKAREKGVAVGQFNINLEWTKAILLTAQENKSPV ILGVSEAGKTYMGTHAVGMVGMGLMDLNIITVEALHLDHSGYEGAFKVIKAEFGSSV MFDGSHYDAENIKETKEVIAKAAKGLSVEAEVGAICGDEEDGVGSGVADPAECKK IADLGVDILAAIGINIKGVYANWAGLRFDVLEDIKNAVGTMPVLVHGTTGIPDDMIA KAIIGLVAKINWNEQCLVPAEATRKVIIEGKQKQGFDPKRLAAGPFAIRAKVKE KMWLFSGVNRA"
gene	complement (9320..10246) /gene="CAC0828" complement (9320..10246) /gene="CAC0828" /codon_start=1 /transl_table=11 /product="Hypothetical protein" /protein_id="AAK78804.1" /db_xref="GI:15023719"
CDS	/translation="MKTIKYLKSNLPFFKKLITLPLFIPLLTGCLNLSKSVTLTSL SKFYNDKAVSINYNPTQWQEGTNRGNQNAIALKKNTDALVGLKCNKSYSENTEDEY ISALNTKYKINSIEGSTSTSVIKVENKKYAVHSRVRADNSIIDSPLVINGDYOE FKPEASDYSKYKYNRDMIMNPKTLKVKYKASPNPTKINIKDLNGSGIKGKNSIT TKRELVLNDSSEKLYKQADERNVITGTYYISKDNILLTILSKVINDGQGVFITASK VIKPKYILKGNTHILNLSISYSIVRLDN"
ORIGIN	Query Match 13.5%; Score 82.4; DB 1; Length 10415; Best Local Similarity 51.4%; Pred. No. 6.6e-14; Matches 216; Conservative 0; Mismatches 201; Indels 3; Gaps 1; QY 100 GGAGCGTCCGGTTCAGGAATAACACTTGGCTCACTGCTTGGCACACTTGCAACCA 159 Db 864 GGTCTAGTGTGTCAGGAATAACCAATAATTAATCTTTTAGTGGATGGATCAAGCA 923 QY 160 AGTTCCCGACAGATCTCTGCGAGATGTAGACCTCTGAAACTCTCTACGCGTAAGCAA 219 Db 924 AGTTCAGCAGTATACGTTGATGGAATGAGATAAGTAAATACATAAAGATGCTA 983 QY 220 CGGTTATACAGGAATAACGTTGGGCTATTATTTCAAGATTATGCTTGGTCCCGAC 279 Db 984 ACGAATACAGCGCGGATGATAGGTTTGTATTTCAATTTTATATCTAGTTCAAAT 1043 QY 280 AGGACAGTTAAATCAACTTCAGCTTGCCTGCGTGGAACACAAATGGCC--TGAATT 336 Db 1044 TTAATCACTAGAAATGAGAACTTGCAGTATGCAAAATCCAAATGGATCCT 1103 QY 337 CCTCACTACTCATGCTGTGCTTGTGCTTGTGCTGCGTTCAGGAAAGCCAGTTTGAATC 396 Db 1104 GTTAAGTGTCTGAAAGCTTTGGCTTAAGATAGGATGATATATTTCTTCCACAGCTT 1163 QY 397 TCTGTCGCGACACACAACTGCTTGGCGGGGTACTGCTCAAAATCCCGAATA 456 Db 1164 TCAGGAGGAGAACACAGAGAGTATCTATAGTAGAGCATTAGCTAAATCCAAAGCTT 1223 QY 457 ATTCTGCTGATCAACCAACCGAGCTCTAGATTAAACAAAGTGTAGCTAGTCATAGAA 516 Db 1224 CTATTGTGTATGAGCTACAGGAGCACTAGATATATACAGGAAGCAATTTAA 1283
RESULT 8	AE013394/c
LOCUS	AE013394
DEFINITION	Methanosarcina mazei strain Goel, section 176 of 379 of the complete genome.
ACCESSION	AE013394
VERSION	AE013394.1
KEYWORDS	GI:20906099
SOURCE	Methanosarcina mazei Goel
ORGANISM	Methanosarcina mazei Goel
REFERENCE	1 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacob, C., Bruggemann, H., Lienard, T., Christmann, A., Boencke, M., Steckel, S., Bhattacharya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. The genome of Methanosarcina mazei: evidence for lateral gene transfer between bacteria and archaea J. Mol. Microbiol. Biotechnol. 4 (4), 453-461 (2002) 22120827 12125824 2 (bases 1 to 12929) Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wierzer, A., Baeumer, S., Jacob, C., Bruggemann, H., Lienard, T., Christmann, A., Boencke, M., Steckel, S., Bhattacharya, A., Lykidis, A., Overbeek, R., Klenk, H.-P., Gunsalus, R.P., Fritz, H.-J. and Gottschalk, G. Direct Submision Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Grisebachstrasse 8, Goettingen 37077, Germany Location/Qualifiers 1..12929 /organism="Methanosarcina mazei Goel" /mol_type="genomic DNA" /strain="Goel" /db_xref="taxon:192952" 333..4805 /gene="cobN" /note="synonym: MM1602" 333..4805 /gene="cobN" /codon_start=1 /transl_table=11 /product="Cobalamin biosynthesis protein CobN" /protein_id="AAK31298.1" /db_xref="GI:20906100" /translation="MLFSSSFAAGSTGTEYELVANTTSABGNVLLSGIPNGNYTLVA ISSDQRRARDEQIQVQGTDLVNLDTSSRDIDETESQAIDLKSGVPLEGASSIRG LTRQMGQDPISLGFTFVLLREQVETPDVSSPGQHLDAIYGYRSHLPLKLM BEINGSNLSLSCFMADYVWENDVLSGMDIITYNNLSPSTARKLTPTVDSASNG CVIIDDILLNESIPLPADLTIESREKLNWYLANGYDQSNKLNLFCTASDCCGRTD LLYDEPHLPERAIYTNWGTCTDLSLDTLTWYNSDGHVVDPAKPTVALIMYQS YFPQLEPIDALISELEARENVAITVGSNLSGDPFKQGDVLDVDAIISFTTFGNK FPAEGLDVPVIGVNTNNRTGYEASRPLPADKMKLDLQELWGAIDPVMATEI DPETGTSVSDYQVWLIDRVESWNLGEIPDSKKVAIYYHGGCKNIGASYL DVFSLSLLDMAAGGVAIENQSLNETELLNLTQGINIGTWAPGELEKMTATGK VVLPESTYVWFSELPEEQVETIQWGPAPGIMVWENESGYLVTPREVGDNVL LAPQSRGLDDSEALVHDSLELPHQVIAFYLMQHSEREGSGSDALVHFGHGTQ EWLFGKGLSEYDMPISLMIGDLPIVYPMYMDGLGEGNEAKERGAVIVHLPIVS AGSGVDYANLSEAHNYGQAEDASLKAARSEVINLTLSLSDDEFNNTAEKNETY FNAVLEGBELDEYKQSNPYGHLIGTSPRGEQLVGMVNSMLGTDFADEVSEFNE TEGSQLYLLDLVNLQGVNSTEAQVQLGTNETVSTFLVSAEYANLAGEEIIQQV LNALDGRFIPGNQGGDPVRNPEALPSGRNFYAFQRIVPTPEAATLGMADQMLEAY LAENGTPRKVAIYMWAGETTRHGVMEATLSLLGVKPVNSGRVVDVEAI PMETLG RPRIDVVIQISGLVDMVDPKVLVLDKAVLAVACEDSPNVI RENADALKTLSSEG LNSEALDLSLRI FEGSEDGAGCTGLPNVSASTWESNDALADLYINMCNAYGENV WGENLKDLFKNLSGVAETVHSRINYLHLDNDFFQYLGGLNLAVSSVSGEYTPDS FITNLGAGDKETLEEFUKRETYISRYFNRPIDWQMGNGYAGARENSDFEVLWGW EATNPDLISDDMNKVTYTMADQELSDWKQNNPYAQSVTFPEATLEATARKGSWASD DVLKSLAEYAEASVYEDGTCHHTCGNSLNNYSLVSGLVSPFGEATIEATQESLKO ADTBHRSSNGHQTSAVKLNKTASNTPEVSNQTVQNSDAGYGVDSPEPAPVRQS ADPYVGYEMQKEESEGAESGMSFGADIIIGTLFVAVVGGIYLGRKKW" complement (4905..5342) /gene="MM1603" complement (4905..5342) /gene="MM1603" /codon_start=1 /transl_table=11 /product="conserved protein" /protein_id="AAK31299.1"

```

/db_xref="GI:20906101"
/translation="MSENENAPEKVEFTPIGYVENDYFKPEYKEVEYKSVILVRE
VLVDGYRVLEKYLIFYSKSEGYKLIHRRYDGMGSGVFAFSRSPYRNPGLTL
VELLVKGVNLYVKGDAINGTPVLDIKPYMKDLSEKENTK"
gene
complement (5335..5853)
/genes="MM1604"
CDS
complement (5335..5853)
/genes="MM1604"
/codon_start=1
/transl_table=11
/product="turgsten formylmethanofuran dehydrogenase,
subunit E"
/protein_id="AA031300.1"
/db_xref="GI:20906102"
/translation="MAVLAERFKORSEDEELVAVENRSCAVDAIQAINGCTCGKN
LIPKEGKVYFYFKGSEKALRISLKPDALQDSRHTALAKRAGTASPEERKEFQ
ASHDASKRILENPESELFRVSEVNIPEPKALIYPTVICKSCGEGFPELGRVKNGE
IVCISCFEAKNE"
gene
7350..8138
/genes="MM1605"
CDS
7350..8138
/genes="MM1605"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AA031301.1"
/db_xref="GI:20906103"
/translation="MNSLLELIFLSEKRNLLFLKEGPKNIEIKTNLDTSVAAIL
PQIKLRDSDLVFRFDSYFLSPLGIAVAERHMAVRVNVFSRNYDYADHAEVCLP
SPUKSRIGLENCISPEPDRWHLPEHREFEKLTGVRKIKGISAIFPEPLPSLPLS
FANGADVLLVTLVPVRIKGEYSELREPLSDNTCLYCEKIKBESFHVTDKPLS
LSLPSNGAFDTRDVLCPDVALKAGDGLFAYDRSEKIKV"
gene
8236..9159
/genes="MM1606"
CDS
8236..9159
/genes="MM1606"
/notes="aIF-2B2"
/codon_start=1
/transl_table=11
/product="Archaeal protein translation initiation factor
2B subunit 2"
/protein_id="AA031302.1"
/db_xref="GI:20906104"
/translation="MKEVQETAEKIRMEIRGAGRIAKAAAEVIRDVAAGLEAASIEE
FAGITVENYLISTRPTAVSNPAVKLSRYSNDVNERKEVINNANLFIIRADKA
LGRIGIGAGRVQDGDVIMTHCNHSAALSITTFEDGKNISVIATSESRPROGLLTV
RLNGPGITPTLLVPSAVRYMKVYVIGADALANGALVNKIGITSOLALAHAR
KSFVAAETFKPSPTILGIPIEIERSTDEVDPVSLAELPHVQVKNFADFPTFSEY
IDMIVDIGIIPANAYIVIKHLEWFEFEI"
gene
9219..9683
/genes="MM1607"
CDS
9219..9683
/genes="MM1607"
/codon_start=1
/transl_table=11
/product="Acetyltransferase"
/protein_id="AA031303.1"
/db_xref="GI:20906105"
/translation="MINVEYTDGAOBSLYLIESLWQSRTHRNKSKYFADTYANKRF
QDRVNLTDSSAAMRVNLVKDQGTQVIGYICSTIDKEMTIGDSLYLVEYKFKHGI
GSLWEKALEWLDKKNKSKIVAVGDGNEKVIDFNYHVGPHIRKILIEQVSN"
gene
complement (9752..10462)
/genes="MM1608"
CDS
complement (9752..10462)
/genes="MM1608"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AA031304.1"
/db_xref="GI:20906106"
/translation="MREKEGETGAGRVIMQIENLSKTYMQCKIPVHALRCACITVR
KGLLALMGPSGSKSTIALIGLTKATDGLDGLTDLTSVPEKLLPVRREKIFG
IQHYNLPTLSALENRFMRWQKKEKERSKLELDEGLDEKHLKHPTELGGG

```

```

EQRVSIARALANKPALILADEPTGEVDSKTRDSVQIFKELSGKQTILVWTHDPEV
AKECRSVLRITDGMKDH"
complement (10455..11570)
/genes="MM1609"
CDS
complement (10455..11570)
/genes="MM1609"
/codon_start=1
/transl_table=11
/product="ABC transporter, permease protein"
/protein_id="AA031305.1"
/db_xref="GI:20906107"
/translation="WLDIIRNITNRKIRGALTICGIALGIFAVIVMGAMSENPHOTF
ERSMVSATDKTRVFAESGVFGGLTEKVSDFLRVAGVEDAYGLLMTPTPDRMGWTG
KQILGVPEKSSVALNPVLKAGFLDPGDSYSVVGNNIKRNVLEIGSKFPIHKY
FTVIGILYTSIFDNVILFLETAQDLISVGNVSVPYPAVDPDRVDAEMLSKRIBLS
VAGTISYFGEVQAQSFMSVITISSGLAAIIGGLCVMTMLMSVAERISFG
ILKAIGATKDLILLTGLSEASLMGLGLGIIVGIGAVQIMNAWLATTRIVLFLTP
RLTIAMFLALLIGALSGLPAYRASKVSPMEALKHA"
gene
complement (11626..12813)
/genes="MM1610"
CDS
complement (11626..12813)
/genes="MM1610"
/codon_start=1
/transl_table=11
/product="hypothetical binding protein of ABC transporter"
/protein_id="AA031306.1"
/db_xref="GI:20906108"
/translation="MSLPKKNCKLREFLSLFLIFFCTOINPVYGVNGAA
LEENENRILIGFSDHLSDATVRPEPHENSVLFTLSSGKETLKSEFHLVSVYK
DOELSKYFWELEEDFKDRSDYTAOLFVKEGNCILETOKTSPSYRNLNKKVDF
SADSEKSFLLDSSANLRLVQMPERGIYDFDLKLLSKDILYSSSLNKPVTDAY
KATNPDLLEKDRSYTALKVHSHPELITAYMSEFWEKEKLEILSDSDINVDYASV
TVTGRSQVFNIGIVRVLTPEEGDVRFEETADITLPGEDVTGIIWQGVPRGDYNVK
IDVNLNDELDSHETVLRVFEVPAELNPEETPALSSQAATGILLCVAVFLGRKKRG
KKD"
ORIGIN
Query Match 13.5%; Score 82; DS 1; Length 12925;
Best Local Similarity 50.8%; Pred. No. 9.3e-14;
Matches 281; Conservative 0; Mismatches 235; Indels 18; Gaps 2;
QY 78 AGGACAAATGACGACACTGAGGCTCGGTTCAGGAAATCGACTTGTCAACTG 137
DB 10328 AGGAGAGCTCTCGCTATAATAGGACCTTCGGGTTCAGGAAATCAACCTTCTGGCTCT 10269
QY 138 TCTTGGCACATTCGACAAACCAAGTTCGGGACAGATCCTTGTGAGGATGTAGACCTTCT 197
DB 10268 CATAGGCACATTTGAAAAGCCCTGACGAAATAATTAATTTCTGACGGGACAGACCTTAC 10209
QY 198 GAAACTCTCTACGGCTAAGCAACCGTTATACAGAAATAATACGGTGGCTATTATTCTCA 257
DB 10208 ATCAGTACCTGAAAAGCTGCTCCCGGGCTGAGAAGAGAAAAATAGGCTTATTATTCTCA 10149
QY 258 AGATTATGCTTGTATCCCGACAGACAGTAAATTCACCTTCAGCTTCGGTGG- --- 313
DB 10148 GCATACAACTGATTCCTACATTTGAGCCCTTGAAAACCTTGAACTTGCATGGCTT 10089
QY 314 -----AAAAACAAAATGCGCTGAAAATTCCTCAAGTACTTCATGCTG- ---TTGG 359
DB 10088 TTCAGAAATGCAAAAAAAGAGAAAAAGGCAAAATCCCTTCTTGAAGACCTGGG 10029
QY 360 TCTTGAGTCGTCGAGGAAAGCCAGTTTTTGAACCTCTCTGTTGGCGAACAACAACGAAC 419
DB 10028 GCTGGAAGACCGGTCAATATAAACCACGAGCTCTCAGGTGGAGAACACGAAGGGT 9969
QY 420 TCGGTTGCCCGGTACTGTCTAAAAATCCCGAAATAATTTCTGGCTGTATGAACCAACCG 479
DB 9968 TTTATTGCAAGGCTCTGCGAAATAAACCCTGCCCTGATTTCTGGCTGACGAGCCTACAG 9909
QY 480 AGCTCTAGATTAAACAACAGTGTAGCTAGTCATTAGAGGATTCAGAGAGCACTGCCGACAA 539
DB 9908 TGAGTGCATAGCAAAACAGGAGACTCCGTTGTGAGATTTTAAAGAAATTGAGCCAGAA 9849

```

Qy	540	AGCGCCACCGTGTGTGTCGTACCGACTCGCC	573
Db	9848	GGGACAGCAATCTGTGTGAATCAGCAAGCC	9915
RESULT 9			
AL596174/c			
LOCUS	96208 bp	DNA	linear
DEFINITION	BCT 06-JUN-2002		
ACCESSION	AL592022		
VERSION	AL596174.1		
KEYWORDS	GI:16415440		
SOURCE	Listeria innocua		
ORGANISM	Listeria innocua		
REFERENCE	1		
AUTHORS	Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dusurget, O., Enian, K. D., Fsihi, H., Portillo, F. G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, I., Hauf, J., Jackson, D., Jones, L. M., Kaerst, U., Kretz, J., Kuhn, M., Kunst, F., Kurapkut, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E., Nedjari, H., Nordisiek, G., Novella, S., de Pablo, B., Perez-Diaz, J. C., Purcell, R., Rammel, B., Rose, M., Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J. A., Voss, H., Wehland, J. and Cossart, P.		
TITLE	Comparative genomics of Listeria species		
JOURNAL	Science 294 (5543), 849-852 (2001)		
MEDLINE	21537279		
PUBMED	11679669		
REFERENCE	2 (bases 1 to 96208)		
AUTHORS	Glaser, P., Frangeul, L. and Rusniok, C.		
TITLE	Direct Submission		
JOURNAL	Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE		
COMMENT	E-mail: pglaser@pasteur.fr		
FEATURES	Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.		
Source	Location/Qualifiers		
1..96208	/organism="Listeria innocua"		
	/mol_type="genomic DNA"		
	/strain="Clp11262"		
	/db_xref="taxon:1642"		
gene	complement(2032..3821)		
	/gene="lin2900"		
CDS	complement(2032..3804)		
	/gene="lin2900"		
	/note="similar to ATP-dependent DNA helicases"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAC98125.1"		
	/db_xref="GI:16415441"		
	/db_xref="GOA:Q92620"		
	/db_xref="SPRMBL:Q92620"		
	/translation="MIEQARILQQNFGYQFRGQIDVTISKLSAGEDTLAIMPTGGG KSLCQIPALLFTIVTVPGLTSLMKQDVALSDGPAITFINSITLREIDRLDA AFSGLKMLYAFERIETPGFRLIEQVLSLFAIDEAGHSIQWGHDFRPSYLSLSDS LDKWTRPLVIALTATAQVSDDICRLLIKPNVITKGFSDNLAFQVVKQDKFL YLIDVLTNKSSTSGIIVASTRKEVERIHSFLKKVSGSMVHGSMTDIAFKQDKFL YDIDRIVATNRAFNGINKSNVRVIVHYNIPRNTIAYQAGRAGRGVPSDCILFPS PQSRIQPLIQSSEDDERKQNEFAKLQMTGIGYTEICLQITVQYFDDEENCGK CNSCLDTRDTTILAQVFCSCIKRMGERFGVLIAKVTGTSADQKVDWRFEELST YGLMKDAQDQVLIDYLTAEKYLQPTDSQPSLKTDAISVLRELKVERQAKR ASKVMQDSSDLDFEKLREVRRELAARHAKHVPVPIIFSDETLEMCAYMPQTDALLVK GIGAMKRDYGAEPFLVQQBAAK"		
	complement(3815..3821)		
RBS	/gene="lin2900"		
	complement(3924..5430)		
gene	/gene="guab"		
	complement(3924..3953)		
terminator	/gene="guab"		

APETLPGDTLLKKKGGKDEMGVNYYSRMTAFNGKDGVGSGKMTTGEKCTSEETGVP
GLYKNTNPPLEKTNWDIDPTGLRIGLRLITNRYKLPIMITENGELGEYDSLTEDHK
IHDEYRIEYIKAHALAEALITDGVEMLGCTCTFTDLLSWLNGYKRYGVFVYDRDE
NDEKELKRYKDSFYWKYKTIENGANLVBEOQG"

gene 9935..10237
CDS /gene="lin2905"
/note="similar to PTS cellobiose-specific enzyme IIB"
/codon_start=1
/transl_table=11
/protein_id="CAC98130.1"
/db_xref="GI:16415446"
/db_xref="GOA:Q926Y5"
/db_xref="SPTREMBL:Q926Y5"
/translation="MKKILVCRAGMSTSLIVTKMAHATSGISEIEIHALPVSSEAS
VDRMDIVMUGPQRYOKPQVDLVOGRIPVIVIDMKDYGMLNGKAVLEKAFABIG"

gene 10258..11623
RBS /gene="lin2906"
CDS /gene="lin2906"
/note="similar to PTS cellobiose-specific enzyme IIC"
/codon_start=1
/transl_table=11
/protein_id="CAC98131.1"
/db_xref="GI:16415447"
/db_xref="GOA:Q926Y4"
/db_xref="SPTREMBL:Q926Y4"
/translation="MSIMSKFEHGMERVLVPVANKNSQRHIAAIRDAFLLVPLIMA
GSIITLINFALVDPGFIKILFLGKIFPNLADQAVFSPVMOGNSIMAILLIFLVA
RNLAIFKQDLDLGLSIGAFITVPTVVDNASTYITKFLGAQSLFVAIIIVAIT
GEVSRARSRLMKIMPDQPPAVARSFKVLIPIIITILFVSINYLIIPIAPEGIL
DLVVTQAPLQDGTNVFVSVIIIGLNSLLWLVGHGNTVAIRDIITEPNLNL
SYVAHQSGAGPVPATWAGLNDGFANGSGMTGLGILLIAFIASRRADYRDIAKSL
APGFENEPVIFGLPVLPIVPIPIIIPAINLIGYFFISTKLIPPVAYQVPWT
FGPLIFPGTGGNLLVGLLCLUAVATVILPVLVSNKIRASDAAMKNASTEQ
"

gene 11635..12519
CDS /gene="lin2907"
/gene="lin2907"
/note="similar to xylose operon regulatory protein and to
glucose kinase"
/codon_start=1
/transl_table=11
/protein_id="CAC98132.1"
/db_xref="GI:16415448"
/db_xref="GOA:Q926Y3"
/db_xref="SPTREMBL:Q926Y3"
/translation="MKIAAFDGTALMGVLPVPHGEIILITKSNEIIASDGDQILAE
KLFLAENTDVTVGIVASPGVYVNPKTGLITMGALRRDNLKLEWLEATGGLPVAIEN
DANCALLAEKWLGGQDLDFCLTIGTIGTGIFSGNALVGRGFRAGFSGYMFSE
PCAFPGKTYLNETTLMVLRQVAILTGRPLKEITGEIEIFANYDAHDPISERLINEF
VTGCTGLNLIYLPDPIHPIFGGITSRPTFTELKHMVASFGLRDIITETATKNQ
AGLLGAVYHLEQENRHE"

gene 12497..12502
RBS /gene="lin2908"
/gene="lin2908"

Query Match 13.3%; Score 80.8; DB 1; Length 96208;
Best Local Similarity 47.2%; Pred. NO. 3.7e-13;
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;
QY 1 ATGATAGAAATCAATGACTCAAGAAATCTTTTGGCGTTCGGATCTTATGCGCAAGGCTC 60
DB 19809 ATGATTGAATTAGTTAATGTCAGTAAAGAACATACAGATAAATGATTTGGAAAAAGTC 19750
QY 61 AGTCATAGTTTATACCAAGGACACATGACACTGACTGAGCGTCCGGTTCAGGAAAA 120
DB 19749 TCTCTTTCTATGGAGCTGGCGAATTTATTCAGTCTCGCGGAGAGTGTAGTGGGAG 19690

QY 121 TCGACTTTTGCTCAACTGTCTTGGGCACACTTGACAAACCAAGTTCGGACAGATCCTTGTG 180
DB 19689 ACAACGCTGCTAAATGTTATCGACACCTAGATTCAAAAGATAGTGGCGCAAGTATTATT 19630
QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACGGGTAGCAACCGGTATATACAGGAAAAATACG 240
DB 19629 AACGAGATGGAATATCAGA-----CGAAAAAAGAGGTTATGACTCTTAAAAAAGAGGTG 19576
QY 241 GTGGGCTATTATTTCAAGATTATGCTTGTGATCCCGACAGACAGATTAAATTCACACCTT 300
DB 19575 TTAGGTTTATATTCCTCAAAATATCTATTGATGGAATGAACAGTGTAGAAAACTTA 19516
QY 301 CAGCTTGGCGTGGAAAAACACAAATGGCTGAAATTCCTCAAGTACTTTCATGCTGTGTTGGT 360
DB 19515 TCCATTACAGGTGGGAAAAATCGCAAGCTGATGATAGACATTTTGAAGAAGTAGGAATG 19456
QY 361 CTGAGTGGTTCGAGAAAAAGCCAGTTTGTGAACTCTCTGGTGGCGACACACACGAACT 420
DB 19455 GATGAGAGCTATTATAGCAAAAAAAGTATACCAATTAAGTGTGGGAGAAAAACACGGATT 19396
QY 421 GCGTTGGCGCGGTACTGCTCAAAAATCCCGAATATCTGGCTGATGAACCAACCGGA 480
DB 19395 GCTATTGTGGCATTTTACTCAACCATTTTCAACTTTTCTGGGACGAACCAACTGGC 19336
QY 481 GCTCTAGATTAAACAAAGTAGAGTAGTCTATAGAACGATTGAGAGCACTCGCGACAAA 540
DB 19335 AATTAGATGATAAAAAACAAAAAATCATTTGAATTTTCTAGCTTGAANAAGCAA 19276
QY 541 GCGGCCACCGTTGTTGTGTCAGCACTCGCCCTCTCCGAGATCAGCGGATACCAATT 600
DB 19275 GGTAGACTATCGTTGTGTCTAGCATACCGGAAATATCAGGAAAAGCAGATCGGCTC 19216
QY 601 AT 602
DB 19215 AT 19214

RESULT 10
AX417049/c
LOCUS AX417049 311208 bp DNA linear PAT 02-SEP-2002
DEFINITION Sequence 4040 from Patent WO228891.
ACCESSION AX417049
VERSION AX417049.1 GI:21449659
KEYWORDS
SOURCE Listeria innocua
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE 1
AUTHORS Kunst, F. and Glaser, P.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 4040 11-APR-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
Location/Qualifiers
1..311208
/organism="Listeria innocua"
/mol_type="unassigned DNA"
/db_xref="taxon:1642"
/note="seq 2058, original length: 3,011,208 replaced
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:
0.900.001 to 1.249.980-seq 4035: 1.200.001 to
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:
1.800.001 to 2.149.980-seq 4038: 2.100.001 to
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:
2.700.001 to 3.049.980"

ORIGIN

Query Match 13.3%; Score 80.8; DB 6; Length 311208;
Best Local Similarity 47.2%; Pred. NO. 5e-13;
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAGAAATCTTTGGCGTTCGGATCTTTATGCGAGTCTTC 60
Db 234809 ATGATGAAATAGTATATGTCAGTAAAGATCAAGATAAATGATTTTGGAAAAAGTC 234750
QY 61 AGTCATAAGTTTTTACAGGAACAATGACGACACTGACTGGAGCGTCGGTTCCAGGAAAA 120
Db 234749 TCTCTTTCTATTGGAGCTGGCGAATTTATTTGCACTCGTCGGGAGAGTGGTAGTGGAG 234690
QY 121 TCGACTTTGCTCACTGCTTTGGCACACTTGACAAACCAAGTTCCGGACAGATCCTTGTC 180
Db 234689 ACAACGCTGCTAAATGTTATCGGACACTAGATTCAAAAGATAGTGGGCAAGTTATTAT 234630
QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACCGCTGAAGCAACGGTTATATACAGGAAAAATACG 240
Db 234629 AACGAGATGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAAAAAAGAGGTG 234576
QY 241 GTGGGCTATTATTTCAGAGTTATGCTTGAATTCGCGACAGGACAGTAAATTCACACTTT 300
Db 234575 TTAGGTTTATATTCGCAAAATTTATCTATTGATGGAATAAGAACAGTGTAGAAAACTTA 234516
QY 301 CAGCTTCGGTGGAAAAACAAATGCGCTGAAATTCCTCAAGTACTTTCATGCTGTTGGT 360
Db 234515 TCCATTACAGTGGGNAATTCGAACTGATGATAGAGATTGGGAAGATGGAATG 234456
QY 361 CTGAGTCTGTCGAGGAAAAAGCCAGTTTTTGAATCTCTGCTGGCGAAACAAACGAACT 420
Db 234455 GATCAGAGCTATTATAGCAAAAAAAGTATACCAATTAAGTGGTGGGAAAAACAAACGATT 234396
QY 421 GCGTTGCGCGGCTACTGCTCAAAATCCCGATTAATTCGCTGATGAACCAACCGA 480
Db 234395 GCTATTGTGCGAATTTTACTCAAAACCAATTTCAACTTTTCTTCGGACGAAACCACTGGC 234336
QY 481 GCTCTAGATTAAACACAGTGTAGTCTATAGAAGCATTTGAGACACTCGCCGCAAAA 540
Db 234335 AATTAGATGATAAAAAACAAACAAAAATCATTTGAATTTTCTAGCTTGAAGAAGCA 234276
QY 541 GCGCCACCGTTGTTGTGTAGTGGCTCGCCCTCTTCGGAGATCAGCGGATACCATT 600
Db 234275 GGTAAAGTACTGCTTTGTGTCACGCTAGCGCCGGAATATCAGGAAAAAGCAGATGGGTC 234216
QY 601 AT 602
Db 234215 AT 234214

RESULT 11
AX413018 349980 bp DNA linear PAT 02-SEP-2002
LOCUS
DEFINITION Sequence 9 from Patent WO0228891.
ACCESSION AX413018
VERSION AX413018.1 GI:21445476
KEYWORDS
SOURCE Listeria innocua
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

REFERENCE
1
AUTHORS Kunst, F. and Glaser, P.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 9 11-APR-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)

FEATURES
source
1. 349980
/organism="Listeria innocua"
/mol_type="unassigned DNA"
/db_xref="taxon:1642"
/note="old seq 9 (684707) splitted in 3-new seq 9 = old
seq 9 from 1 to 349980-seq 4027 = old seq 9 from 300000 to
649980-seq 4028 = old seq 9 from 600000 to 684707"

ORIGIN

Query Match 13.3%; Score 80.8; DB 6; Length 349980;
Best Local Similarity 47.2%; Pred. No. 5.1e-13;

Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;
QY 1 ATGATAGAAATCAATGACCTCAGAAATCTTTGGCGTTCGGATCTTTATGCGAGTCTTC 60
Db 346005 ATGATGAAATAGTATATGTCAGTAAAGATCAAGATAAATGATTTTGGAAAAAGTC 346064
QY 61 AGTCATAAGTTTTTACAGGAACAATGACGACACTGACTGGAGCGTCGGTTCCAGGAAAA 120
Db 346065 TCTCTTTCTATTGGAGCTGGCGAATTTATTTGCACTCGTCGGGAGAGTGGTAGTGGAG 346124
QY 121 TCGACTTTGCTCACTGCTTTGGCACACTTGACAAACCAAGTTCCGGACAGATCCTTGTC 180
Db 346125 ACAACGCTGCTAAATGTTATCGGACACTAGATTCAAAAGATAGTGGGCAAGTTATTAT 346184
QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACCGCTGAAGCAACGGTTATATACAGGAAAAATACG 240
Db 346185 AACGAGATGGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAAAAAAGAGGTG 346238
QY 241 GTGGGCTATTATTTCAGAGTTATGCTTGAATTCGCGACAGGACAGTTAAATTCACACTTT 300
Db 346239 TTAGGTTTATATTCGCAAAATTTATCTATTGATGGAATAAGAACAGTGTAGAAAACTTA 346298
QY 301 CAGCTTCGGTGGAAAAACAAATGCGCTGAAATTCCTCAAGTACTTTCATGCTGTTGGT 360
Db 346299 TCCATTACAGTGGGNAATTCGAACTGATGATAGAGATTGGGAAGATGGAATG 346358
QY 361 CTGAGTCTGTCGAGGAAAAAGCCAGTTTTTGAATCTCTGCTGGCGAAACAAACGAACT 420
Db 346359 GATCAGAGCTATTATAGCAAAAAAAGTATACCAATTAAGTGGTGGGAAAAACAAACGATT 346418
QY 421 GCGTTGCGCGGCTACTGCTCAAAATCCCGAATTAATTCGCTGATGAACCAACCGGA 480
Db 346419 GCTATTGTGCGAATTTTACTCAAAACCAATTTCAACTTTTCTTCGGACGAAACCACTGGC 346478
QY 481 GCTCTAGATTAAACACAGTGTAGTCTATAGAAGCATTTGAGACACTCGCCGCAAAA 540
Db 346479 AATTAGATGATAAAAAACAAACAAAAATCATTTGAATTTTCTAGCTTGAAGAAGCA 346538
QY 541 GCGCCACCGTTGTTGTGTAGTGGCTCGCCCTCTTCGGAGATCAGCGGATACCATT 600
Db 346539 GGTAAAGTACTGCTTTGTGTCACGCTAGCGCCGGAATATCAGGAAAAAGCAGATCGGTC 346598
QY 601 AT 602
Db 346599 AT 346600

RESULT 12
AX417036 349980 bp DNA linear PAT 02-SEP-2002
LOCUS
DEFINITION Sequence 4027 from Patent WO0228891.
ACCESSION AX417036
VERSION AX417036.1 GI:21449646
KEYWORDS
SOURCE Listeria innocua
ORGANISM Listeria innocua
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

REFERENCE
1
AUTHORS Kunst, F. and Glaser, P.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 4027 11-APR-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)

FEATURES
source
1. 349980
/organism="Listeria innocua"
/mol_type="unassigned DNA"
/db_xref="taxon:1642"
/note="seq 9 splitted in 3 part.-seq 9: 1 to 349980-seq
4027: 300001 to 649980-seq 4028: 600001 to 684707"

ORIGIN

Query Match 13.3%; Score 80.8; DB 6; Length 349980;

Best Local Similarity 47.2%; Pred. No. 5.1e-13; Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;		Jouy en Josas 78352, France Location/Qualifiers 1. 11551	
Qy	1	ATGATAGAAATCATGACCTCAAGAATCTTTGGCGTTCCGATCTATATGGCAAGTCTC 60	/organism="Lactococcus lactis subsp. lactis"
Db	46005	ATGATAGAAATCATGATGCTAAAGAAATCAAGATAAATGATTTGGAAAAGTC 46064	/mol_type="genomic DNA"
Qy	61	AGTCATAAGTTTTTACCAGGAACAATGACAGCACTGACTGAGCGTCCGGTTCCAGAAA 120	/strain="IL1403"
Db	46065	TCCTTTCTATGAGCTGGCGAATTTATTCAGTCTGCGGAGAGTGGTAGTGGCAAG 46124	/sub_species="lactis"
Qy	121	TCGACTTTGCTCAACTGCTTGGCACACTTGCACAAACCAAGTTCGGACAGATCTTTGTC 180	/db_xref="taxon:1360"
Db	46125	ACAACGCTGCTAAATGTTATCGGACACCTAGATTCAAAAGATAGTGGCGAAGTTATTAT 46184	86. 328
Qy	181	GAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCAACGGTTATACAGGAAAAATACG 240	/gene="ylac"
Db	46185	AACGAGATGGAATACGA-----CGAAAAAGAGGTTATGACTCTAAAAAAGAGGTG 46238	/note="synonym: L104681"
Qy	241	GTGGCTATTATTCAAGATTATGCTTCCGACAGGACAGTTAAATTCACCTT 300	96. 328
Db	46239	TTAGGTTTATATTCCAAATTTATCTATTCATGGAATAATGAAACAGTGTTAGAAAACCTTA 46298	/gene="ylac"
Qy	301	CAGCTTCGGTGGAAACAAATGCGCTGAATCTCTCAAGTACTTCAATGCTGTGGT 360	/note="synonym: L105113"
Db	46299	TCCATTACAGTGGGGAATTCGCAAGCTGATGATAGAGCATTTGGAGAGTAGGAATG 46358	539. 1750
Qy	361	CTTGAGTGGTTTCGAGGAAACCCAGTTTTTGAATCTCTCTGTCGCGAACCAACGAAC 420	/gene="ylad"
Db	46359	GATGAGAGCTATTTAGCAAAAAGATATACCAATTAAGTGGTGGAGAAAACCAACGGATT 46418	/note="EVIDENCE BY HOMOLGY BIO15.01 HYPOTHETICAL"
Qy	421	CGTTGGCCGGTACTGCTCAAAAATCCCGGAATATTCGCTGATGAACCAACCGGA 480	539. 1750
Db	46419	GCTATTGGCCATTTTACTCAAAACCAATTCACATTTTCTGCGGACGAACTGGC 46478	/gene="ylad"
Qy	481	GCTCTAGATTTAACAAACAGTGAGCTAGTATAGAACATTTAGAGCACTTCGCGGACAAA 540	/note="EVIDENCE BY HOMOLGY BIO15.01 HYPOTHETICAL"
Db	46479	AAATTAGATGATAAAACAAACAAAATCATTTGATTTTCTAGCTTTGAAAAGCAA 46538	539. 1750
Qy	541	GGCGCCACCGTTGTTGTTGCTACGCACTCGCCCTCTTCGAGAAATCAGCGGATACCAIT 600	/gene="ylad"
Db	46539	GGTAAGACTATCGTTGTGTGTCACGCATGACCCGGAATATACGAAAAGCAGATCGGTC 46598	/note="synonym: L106374"
Qy	601	AT 602	complement (1773. 2360)
Db	46599	AT 46600	/gene="ylac"
RESULT 13 AE006342		11551 bp DNA linear BCT 14-MAY-2001	
LOCUS		Lactococcus lactis subsp. lactis IL1403 section 104 of 218 of the	
DEFINITION		complete genome	
ACCESSION		AE006342 AE005176	
VERSION		AE006342.1 GI:12724053	
KEYWORDS		Lactococcus lactis subsp. lactis	
SOURCE		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
ORGANISM		Lactococcus	
REFERENCE		1 (bases 1 to 11551)	
AUTHORS		Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarme,K.,	
TITLE		The complete genome sequence of the lactic acid bacterium	
JOURNAL		Lactococcus lactis ssp. lactis IL1403	
MEDLINE		Genome Res. 11 (5), 731-753 (2001)	
PUBMED		21235186	
REFERENCE		2 (bases 1 to 11551)	
AUTHORS		Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarme,K.,	
TITLE		Weissenbach,J., Ehrlich,S.D. and Sorokin,A.	
JOURNAL		Direct Submission	
MEDLINE		Submitted (09-JAN-2001) INRA, Genetique Microbienne, Domaine de	
PUBMED			

Jouy en Josas 78352, France Location/Qualifiers 1. 11551		Lactococcus lactis subsp. lactis	
/organism="Lactococcus lactis subsp. lactis"		/mol_type="genomic DNA"	
/strain="IL1403"		/sub_species="lactis"	
/db_xref="taxon:1360"		86. 328	
/gene="ylac"		/note="synonym: L104681"	
96. 328		/gene="ylac"	
/note="EVIDENCE BY HOMOLGY BIO15.01 HYPOTHETICAL"		General	
/codon_start=1		/transl_table=1	
/product="HYPOTHETICAL PROTEIN"		/protein_id="AAK05195.1"	
/db_xref="GI:12724054"		/translation="MAITNEQVERINELARKKKAEGLEAELEFEQALLRRAYLDSVKA	
NFRSQVETIKVIDEKTGEDVTPDKKEIQKNGNRD"		539. 1750	
/gene="ylad"		/note="synonym: L105113"	
539. 1750		/gene="ylad"	
/note="EVIDENCE BY HOMOLGY BIO15.01 HYPOTHETICAL"		General	
/codon_start=1		/transl_table=1	
/product="HYPOTHETICAL PROTEIN"		/protein_id="AAK05196.1"	
/db_xref="GI:12724055"		/translation="WNLKVGISRSAILYVFLAVIIGLIGIVDAIFGRGLIYITEPR	
DSHALYLPFTLAGLIVYIYKNGKDSOKGMLIPEANGHREDIPKRLVPLIVES		2360	
TWLHLFGASVREGVAQVGGVIGHAIGKLSAKEAKILLITGMAAGFAGLEQTP		/gene="ylac"	
AAATFAIEILMLGKIEYALIPALVSVASWTSSSISGLEKFSFAMINIHIDPVL		/note="synonym: L106374"	
KLAIVAVCGIVGFRFAESLAFMKATVAKIVNPFYRIILMGIVISIGLIVHLDRVA		complement (1773. 2360)	
GLGNLISLSPNGGHINGDWIKLIIFTVLSISAGFQGVETPLFAIGSTLGAALML		/gene="ylac"	
FGLPVEFAAGVYSIFSAATNSYFGFIFAAEVFGFSGVQYILPIMTIAYVLNGSS		/note="EVIDENCE BY HOMOLGY BIO15.01 HYPOTHETICAL"	
IYAEVLNLEV"		General	
/codon_start=1		/transl_table=1	
/product="HYPOTHETICAL PROTEIN"		/protein_id="AAK05197.1"	
/db_xref="GI:12724056"		/translation="MDKGRKKVGLGNGFNFIHHAHLMADQVAQOMNLDKVLIMPE	
NIPHVDEKETSIAKHVKMLLEAIKENPRLGLELIERGKGSYSDTLKLTPEANP		2360	
DTDYFIIIGSDMVVLPKWVKIDELIKLIVTFLAIRRKDTISKSPYVTVLDAPLLPIS		/gene="ylac"	
STMLKMPAKNIEITYFLPQNVIDYIKTEKLYKR"		/note="synonym: L107468"	
2885. 4372		/gene="ylac"	
/note="EVIDENCE BY HOMOLGY BIO15.02 HYPOTHETICAL"		General	
/codon_start=1		/transl_table=1	
/product="conserved hypothetical protein"		/protein_id="AAK05198.1"	
/db_xref="GI:12724057"		/translation="MGEKIMPHLYODDSLTLHTDLYOINMOTYFELGRHORQAVFEV	
FREMPPKNGAVFAGLEMYDYLNKLFSETDIAYRELDYPEDFLNLYKELRFTGT		2360	
VRAMGEQALFANPEFLVVEGSLAECCQLVETALININFTQLIATKARIKIVIGDOP		/gene="ylac"	
LLEFCTRAQEAADAAINGADATSNVRAAKIFGUKASGTHAHUUVQSYGND		/note="synonym: L107468"	

YEAFAKAYATHKDCVPLVDTYDLRVGPSALVAREFGDKINFGVRIDSGDMAYIS
KKVROOLDAGSTQDAKIYASNDLDENTILSLAKQAKAIDVWGVTGKLLITAYDQPALGA
VYKLSMEDSESGMDQTIKLSNAKVSPTGKQKQWRLTKKADGSEGDYITFSDERP
DVAREIYMRHPDYTYINKNVRPNANPLKDI.FVGGELIYGLPSLQEIKEFARENLEE
LWDEYKRDNDYFYPVDDLSQELNWHKRLISQVRTVSNTRLDDDE"
/gene="yiaG"
/notes="synonym: L108967"
4369. .4857
/gene="yiaG"
/note="EVIDENCE BY HOMOLGY BIO15.01 HYPOTHETICAL"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05199.1"
/db_xref="GI:12724058"
/translation="MMRIERIDLCDARLENDYQKLDKTESKFMLEPERBERDFRRTI
GRISRPDLVAGDVCKFVGFSLAKGNANRINRAYVYVGIIEBPFQHGIGSQLPFE
LWKAKEKAVKRLLELIVMINNVAIALFQFHGFTIEGVRRVAVCVDDKRIIDFYMSKI
LD"
4918. .5742
/gene="nadE"
/notes="synonym: L0203"
4918. .5742
/gene="nadE"
/notes="EVIDENCE BY HOMOLGY BIO02.12 Pyridine nucleotides.
DE NOVO BIOSYNTHESIS OF NAD"
/codon_start=1
/transl_table=11
/product="NAD-synthetase (EC 6.3.5.1)"
/protein_id="AAK05200.1"
/db_xref="GI:12724059"
/translation="MTLQBEIKELGVKVIDKEIRVSVFLKDYLLKKYPIKSFV
LGISGGQDSLAGRLAQIAIEEMQDSQADETVYFAIRLIFYGVQADEEDAQALAFIQ
PDVSLVNTVKAAYGVQALNEAGIEVDFNKNIRKARQMITQAVAGQYQGVAVLGT
DHAENITGFTFFKGQGGADLLPLFLNKRQKALLAELGADPAIYKVPYTDLEEGK
PGLADIALGVYNDIDYTEGVISEDAKAKIEMWKKTKQKRLHLPISVDFDFWK"
5870. .6571
/gene="yiba"
/notes="synonym: L110431"
5870. .6571
/gene="yiba"
/notes="EVIDENCE BY HOMOLGY BIO13.01 TRANSPORT AND BINDING
PROTEINS General. There are 60 paralogs in L.lactis."
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="AAK05201.1"
/db_xref="GI:12724060"
/translation="MTPIEVNESKIYQSGEKIYANDOVNPEIEKGLAVILGSSGA
KSTVNLILGMDTNDGEVI IDEBKLSNFKELITYRVAVGFQFQVANNLITGA
LENVELASEIVENALDAKEVLKSVGLEHRLHNPFSQISGGEQORVALAALAKQPKIL
LCDEPTGALDYHTGKQKILQIDMARKEGRTVIIVTHNMAIAFIANRVIQMDHGKIQ
IDNNSQPEDIDRIEW"
6565. .9255
/gene="yiba"
/note="synonym: L111162"
6565. .9255
/gene="yiba"
/note="EVIDENCE BY HOMOLGY BIO13.01 TRANSPORT AND BINDING
PROTEINS General"
/codon_start=1
/transl_table=11
/product="ABC transporter permease protein"
/protein_id="AAK05202.1"
/db_xref="GI:12724061"
/translation="MWSYAKKTLNKEIRITSGLSRPFISFILMLLGTFAFVLKLVIS
GPNMRRTAEDPYAGHLADLTLTSLGLSDQDLINETKGVKKAEGFYQDILVIGK
ENSLRFLSKPGLSTYELMSKLPQKDEIADLYDQYKIQGTIDFTPPKSKDGL
IKNHSFKIVGVKSVVDKDFGFTVGTGKLVGVALVTKFAFDSVDVYMIARLSYKN
LQNTLIPDSVSLKTEKLTENTFKNQPEKRLAALKTGPEKQINEAKSOIIVEENQ
LTQENQLIAQNKQIGNASQAQLEQINACQNNQNDGKEXIAKAKELAKQETALQOL
EKPTYIQNNLKEGPNFGYKTLDDSDTRIDSLSNIFPVVILFAILLVSLTWTNTRFVEER

GNLGLLKALGYSNDRIRKFMVYGLVSSGALVGLTIIIGHTFLPIAVFNAYTASSTFS
NRLTSLPTIIVAFALAIACSLIPAYVVMELKEVPASFLAKPKAGSRILLLEKI
NFIWKMSFTYKVTARNLRYFKRMLMTIPVAGCTALLVMGFIKRSISGLSNKQFG
QILHYDMITTEKKNVDKEEEDKELASSEIENYLPIAFENLTKEASGLKEQVNL
IATNRSDDLKSVTSLSKRNQSKTELNSGAVSELSEFAELDLKVGSLTLKQSENP
HKIKAAALTEMNGHYIPNQSVYQKFKDFVANAQLIKLKNASTNNIKMSAQFQ
EDGICVQNSDLKVKVIESMYGLNSVNLISCAILGALIVWLNUNINVSRIEEL
ATIKVLGFYDREVTLYIRETILLISFLGILVFGGLGDFYHQVIMNQLSADQIMPAPGL
LWNLILLSAATFATTTLLAIVVHFLKAVDMLGALKSVDP"
complement (9331. .10116)
/gene="cobQ"
/note="synonym: L113931"
complement (9331. .10116)
/gene="cobQ"
/note="EVIDENCE BY HOMOLGY BIO02.09 Riboflavin and
cobalamin. 43% identical to cobyric acid synthase CobQ of
Heliobacillus mobilis, cobalamin biosynthesis pathway"
/codon_start=1
/transl_table=11
/product="cobyric acid synthase"
/protein_id="AAK05203.1"
/db_xref="GI:12724062"
/translation="MAYISLKSLENPKYSLNVAHLYGDLMTYTGDNINMLKYVGE
KLGAEMTFNIVSLGDTYPAKEDYDLVFWGGQDYEOEIIADQSLIDLAPLKDYTESK
PLLAICGGYQMLGOVYVNSEGTQKIQGTILGHVYENLRTDRFIGDIETHNEEFQETYY
GFENHSGITVLSLDEKPLGTVVVGGNNPDQTEGLYKNTFGYFHPILSRNARLA
YRLVTIAYKIGIEIQLPAPEEILADEDKQQIGDLKRVKVK"
complement (10117. .11466)
/gene="yibD"
/note="synonym: L114717"
complement (10117. .11466)
Query Match 13.2%; Score 80.6; DB 1; Length 11551;
Best Local Similarity 48.9%; Pred. No. 2.5e-13;
Matches 246; Conservative 0; Mismatches 254; Indels 3; Gaps 1;
QY 59 TCAGTCATAAGTGTTTTACAGGACAATGACAGCAGCTGACTGGAGCGTCGCGTT CAGGAA 118
DB 5946 TAAATTTTGAATTTGAAAAGAGAGAAATTGGCTGTCTATTTTAGTTCTCTCGGGCAGGTA 6005
QY 119 AATGACCTTGTCTCACTGTCTTGGCACACTTGACAAACCAAGTTCGGACAGATCCCTTG 178
DB 6006 AATCAACAGTCTCTAATAATTTTAGGAGGAATGGATACAAATGATGAGGAGAGGTTATTA 6065
QY 179 TCGAGGATGTAGACCTTCTGAAACTCTCTACCGCTAGCAACCGTTATACAGGAAAAATA 238
DB 6066 TTGATGAGAAAAAATCTCTAATTTTCAACAAGAATTCATTACCTACCGAGCTTATG 6125
QY 239 CGGTGGGTATTATTTCAAGATATATGCTTGTATCCCGACAGACAGTAAATTCACAC 298
DB 6126 CGGTGGTTCGTTTTTCAATTTTATATCTTGTCAATAATTGACAGCAGCTAGAAAAATG 6185
QY 299 TTCAGCTTCGGTGGAAAAACACAAATGCGCTGAAATTCCTC--AAGTACTTTCATGCTG 355
DB 6186 TCGAATTAGCTTCTGAAATTTGCGAAATGCTCTTGATGCTTAAGAAGTGTCTGAGAGTG 6245
QY 356 TTGGTCTTGAAGTCTGTCGAGGAAAAAGCCAGTTTTTGAACCTCTCTGGTGGCAACCAAC 415
DB 6246 TCGGTCTGGAACATCGGCTTCATAATTTTCTTCGCAACTTCTGCTGGCAGCAGCAAA 6305
QY 416 GAACTGGTTCGCCCGGCTACTGCTCAAAATCCCGAATATCTCTGCTGATGAACCAA 475
DB 6306 GAGTAGCAGATTGCGGTGCCCATAGCCAAAAATCCAAAAATCTTGCTTTGTGATGAACCAA 6365
QY 476 CCGGAGCTCTAGATTTTAAACAACAGTGTAGTGTATGAGAAGCATTTGAGACACTTCGCCG 535
DB 6366 CCGGTGCCTTGATTTATCATACAGAAAAACAATCTTGAAAAATCTTCAAGATATGGCCA 6425
QY 536 ACAAGGCGCCACCGTGTGTGT 558
DB 6426 GAAAAAGAGGAAAAACAGTTATT 6448


```
RESULT 14
SCU96166
LOCUS
DEFINITION
  Streptococcus crista 19841 bp DNA linear BCT 03-APR-2000
  ATP-binding cassette transporter-like protein (tptA), ATP-binding
  cassette protein (tptC) and ATP-binding cassette transporter-like
  protein (tptD) genes, complete cds and unknown genes.
ACCESSION
  U96166 AF227987
VERSION
  U96166.2 GI:7363474
KEYWORDS
  .
SOURCE
  Streptococcus cristatus
  Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
  Streptococcus.
REFERENCE
  1 (bases 1 to 4396)
  Correia, F.F., Lamont, R., Bayer, M., Rosan, B. and DiRenzo, J.M.
  Cloning and Sequencing of a Mutated Locus that Affects Fibrial
  Tuft Organization and Cornicob Formation in Streptococcus crista
  CC3A
JOURNAL
  Int. J. Oral Biol. 22 (4), 241-248 (1997)
REFERENCE
  2 (bases 4397 to 19841)
  Correia, F.F., Allen, T.W. and DiRenzo, J.M.
  High molecular weight serine-rich protein gene (srpA) from
  Streptococcus crista
  Unpublished
JOURNAL
  3 (bases 1 to 4396)
  Correia, F.F. and DiRenzo, J.M.
  Direct Submission
  Submitted (02-APR-1997) Microbiology, University of Pennsylvania,
  4010 Locust, Philadelphia, PA 19104-6002, USA
REFERENCE
  4 (bases 1 to 19841)
  Correia, F.F., Allen, T.W. and DiRenzo, J.M.
  Direct Submission
  Submitted (24-MAR-2000) Microbiology, University of Pennsylvania,
  4010 Locust, Philadelphia, PA 19104-6002, USA
REMARK
  Sequence update by submitter
COMMENT
  On or before Mar 31, 2000 this sequence version replaced
  gi:6984159, gi:2822196
FEATURES
  Location/Qualifiers
    1..19841
      /organism="Streptococcus cristatus"
      /mol_type="genomic DNA"
      /strain="CC3A"
      /db_xref="taxon:45634"
      /note="B15 locus"
    625..870
      /gene="tptA"
    625..870
      /gene="tptA"
    625..870
      /gene="tptA"
    site of signal peptidase II"
    /codon_start=1
    /transl_table=11
    /product="ATP-binding cassette lipoprotein"
    /protein_id="AA897959.1"
    /db_xref="GI:2822197"
    /translation="MLVLLGMSGVFFLAACGSKNADNKEISDGKVDKFKDAVDQKTK
    TVDGKVTETMPDGNVIQIPADGELPTDESESTDK"
    899..2164
      /gene="tptB"
    899..2164
      /gene="tptB"
    /note="tptB"
    /codon_start=1
    /transl_table=11
    /product="ATP-binding cassette transporter-like protein"
    /protein_id="AA897960.1"
    /db_xref="GI:2822198"
    /translation="MFQIKKNNKTKILLGSTVILACAGLGGVYLLHQNQQQAMKRV
    NKDSLTVQADVANNKTSILVLSGEVANNSSKVIDPSKGEYKVFVNGDVTQGG
    PLFSYVLTQASQAYDAQAKANSITAAQTASIKWTYNNKLANLNTLNKYNSS
    KDSLLDOIKSADSLAQSLSDAKTADNEVTNAQIEAKAQVTAOTESDRMKYDVTIA
    DTAGTITSNEDLPTQSKAKKEETIEIMDKSKTLIKGSVSEFDRKLSVQRVDV
    repeat_region
  repeat_region
    complement(6301..16446)
    /note="ORF9R; high molecular weight, serine-rich protein;
    contains cell wall anchor motif for gram-positive cocci
    cell surface protein; contains two amino acid repeat
    regions"
    /codon_start=1
    /transl_table=11
    /product="srpA"
    /protein_id="AAF34780.1"
    /db_xref="GI:6984160"
    /translation="MVMRLTKGNWIKERNMFRSKFNKDFIVDEPKYKLYKSGK
    NWKASQSLDLFRIGMGGEVSVSTSDTELDHAYLNFNTLAGIGIILAAGAFVM
    TQGVTVADTSSEYTKVIAEPKAAETSAHSIIAQAEVKKAKAQASATASV
    SVSGVEDQASQSQSVSVSASQASQASQASQASQASQASQASQASQASQASQASQ
    DRKDKRWSGTGTQVGTLLTNTGNSNGKNQOENPNQKGFPTVVELDQGEPLVG
    SHSYVNVENAPEAGKVVVNTKAYTFKNGKTYVWKVEGKVKYKMEVTKKVSDDLVEI
    TEGLTAKDTISTPREGMDGMEVGQNVKA"
    2151..2825
      /gene="tptC"
    2151..2825
      /gene="tptC"
    /note="tptC"
    /codon_start=1
    /transl_table=11
    /product="ATP-binding cassette protein"
    /protein_id="AA897961.1"
    /db_xref="GI:2822199"
    /translation="MLKLNHKSYQQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ
    STLNNIIGCLDKASAGSYHIEGTVDLSNQLSDLRNKKIGFQFNFLMPKLTACQ
    NVELPLTYMKVKERRERALEMLRLVGLERSDFKPMELSGGQKQVIAIALVTPN
    SPILGDEPTGALDTKTSQIMELPKQFNEQKTVIITHEPEVAQLCKQTVVLRDGI
    ETRALG"
    2884..3888
      /gene="tptD"
    2884..3888
      /gene="tptD"
    /note="tptD"
    /codon_start=1
    /transl_table=11
    /product="ATP-binding cassette transporter-like protein"
    /protein_id="AA897962.1"
    /db_xref="GI:2822200"
    /translation="WRSQALMGIIGIGIAIIAIFSIIEGNTENTKQLIGSGNNTIK
    VVYKSAIDPSIPEKSAOKPSYIPWGEDVLSKKEIPIQVKNALTYGADKIVYL
    SKSKSKVQAVSQSVADIKQORLLEGEGFSEAFKQEOVAYLEKSLYDLILFPKGDGI
    KRYVEKGNPKFYGVFESTEGSGSKVATIPQQRHIFDTLWSEFVTVQHL
    KADLLKVAKKYDYLNOQPMQSDYMFGLNLQEFEROLNQLNQSFLVLAGIASL
    LVGIGVNNIMLVSVTERTRIEIGKALGARRKILASGRLPSVSNFSSLGNTSTSR
    4765..5103
      /note="ORF7; similar to transposases"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAF61313.1"
    /db_xref="GI:7363475"
    /translation="MWDAKRRVLKSLKHGIIODKRGNVWRAVFGMDVSKTSSEVAI
    LVNGEKVHYTILNDAIGFNRLGLDKTVHNPEIPEATGVYSRRLQAFLEESGYAI
    PLAIQSQASS"
    5225..6025
      /note="ORF8"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAF61314.1"
    /db_xref="GI:7363476"
    /translation="WIEDLVAKNRLHKVQLVTFPELENLLSTPTGEQYNLVIAPFC
    KEFVLISQSLYEIIHQSKYSRISERKIAIYTDKLKAKQSFCAVKTSPLMEVVR
    YTAQELLRLSERQVNLMDVALQAQPLPEYILRSPIGAETATSIIGELGDIHRFQ
    SYQNFAPIGIDILHVESRNFPAKEHITKRGNYARKILFKFCIHNTASHTNPFCHIA
    DFVEKRRQSTASTKELTIATSHRLIRTWYLYLTHNKLYDSLTQNR"
    5471..6202
      /rpt_type=direct
    complement(6301..16446)
    /note="ORF9R; high molecular weight, serine-rich protein;
    contains cell wall anchor motif for gram-positive cocci
    cell surface protein; contains two amino acid repeat
    regions"
    /codon_start=1
    /transl_table=11
    /product="srpA"
    /protein_id="AAF34780.1"
    /db_xref="GI:6984160"
    /translation="MVMRLTKGNWIKERNMFRSKFNKDFIVDEPKYKLYKSGK
    NWKASQSLDLFRIGMGGEVSVSTSDTELDHAYLNFNTLAGIGIILAAGAFVM
    TQGVTVADTSSEYTKVIAEPKAAETSAHSIIAQAEVKKAKAQASATASV
    SVSGVEDQASQSQSVSVSASQASQASQASQASQASQASQASQASQASQASQASQ
```


[illegible]

QY	270	GATTCCCGACAGACAGTAAATTCACCTTCAGCTTGCGGTGGAAAAACACAATAATGGCC	329
Db	2432	CATGCGCTAAGCTGACAGCCTGTCAAATGTTGAATGCCTTCACATTACATGAAGTTC	2491
QY	330	TGAAT-----TCTCAAGPACTTCATGCTGTGGTCTTGAGTGGT	371
Db	2492	TAAAAGAAGAGAGACGCGAGCGCGCTTTGGAATGCTGCCGATTCGCTCGGTCTGGAAGAGAG	2551
QY	372	CGAGGAAAAGCCAGTTTTTGAACCTCTCTGGTGGCGAACACAAACGAGACTCGTTGGCCCG	431
Db	2552	AAGCGACTTTAAGCCCTATGAGCTGTCCGCGGTCAAAGCAACGGGTAGCTATTGCGCG	2611
QY	432	GGTACTGCTCAAAAATCCCCGMAATAATTCCTGGCTGATGAACCAACCGAGCTCTAGATTT	491
Db	2612	TGCTCTGGTGACCAATCCCAGCTTTATTCGGGAGCAGGACCGACTTGAGCGCTGGATAC	2671
QY	492	AACAAACAGTAGCTAGTCTATAGAACGANTTGAGAGCACTCGCCGACAAAGCGCCACCCT	551
Db	2672	CAAGACAGTGTTCAGATCATGCGATTGTTTCAAGCAGTTTAACGAACAAGGCAAAACTAT	2731
QY	552	TGTTGTGTCTACGCACTCGCC	572
Db	2732	TGTCATCATCCACGAGCC	2752

RESULT 15	AE006565/c	10738 bp	DNA	linear	BCT 01-JUN-2001
LOCUS	Streptococcus pyogenes M1 GAS strain SF370, section 94 of 167 of the complete genome.				
ACCESSION	AE006565 AE004092				
VERSION	AE006565.1 GI:13622379				
KEYWORDS					
SOURCE	Streptococcus pyogenes M1 GAS				
ORGANISM	Streptococcus pyogenes M1 GAS				
REFERENCE	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
AUTHORS	1 (bases 1 to 10738)				
	Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Surorov,A.N., Kanton,S., Qian,Y., Jia H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J., Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.				
TITLE	Complete genome sequence of an M1 strain of Streptococcus pyogenes				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)				
MEDLINE	21192684				
PUBMED	11296296				
REFERENCE	2 (bases 1 to 10738)				
AUTHORS	Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Surorov,A.N., Kanton,S., Qian,Y., Jia H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J., Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-APR-2001) Department of Microbiology and Immunology,				
	University of Oklahoma Health Sciences Center, 940 SL Young Blvd,				
	Oklahoma City, OK 73104, USA				
FEATURES	Location/Qualifiers				
source	1..10738				
	/organism="Streptococcus pyogenes M1 GAS"				
	/mol_type="genomic DNA"				
	/strain="SF370"				
	/serotype="M1"				
	/db_xref="taxon:160490"				
	complement(89..1429)				
	/gene="SPY1252"				
	complement(89..1429)				
	/gene="SPY1252"				
gene	/note="Best Blast hit = sp p47394 Y148_MYCGE_HYPOTHETICAL PROTEIN MG148 >gi 1361569 pir D64216 hypothetical protein homolog MG148 - Mycoplasma genitalium				
CDS	>gi 3844741 gb AACF71366.1 (U39695) conserved hypothetical protein [Mycoplasma genitalium]"				
	/codon_start=1				
	/transl_table=11				


```

gene      complement(8639..8839)
          /gene="SPY1261"
CDS       complement(8639..8839)
          /gene="SPY1261"
          /note="Best Blastp hit = emb|CA885690.1| (AJ271696)
          mismatched base pair and cruciform DNA recognition protein
          [Agaricus bisporus]"
          /codon_start=1

Query Match      13.1%; Score 80; DB 1; Length 10738;
Best Local Similarity 49.0%; Pred. No. 3.7e-13;
Matches 243; Conservative 0; Mismatches 250; Indels 3; Gaps 1;

QY 66  TAAGTTTTTACCAGAACAACTGACACTGAGCGTCCGGTTCAGGAAATCGAC 125
Db 6450  TGAGATTATAAAGAGAACTAGTGGTTATTTGGGAGATCGGAGCTGGTAAATCGAC 6391
QY 126  TTGCTCAACTGCTTGGCACATTGACAAACCAAGTTCGGACAGATCCTTGTGAGGA 185
Db 6390  AGTCCTCAATATTTTGGGGGCGATGGACACAGTAGATGCAAGTCAAGTAATTATTGATGG 6331
QY 186  TGTAGACCTTCTGAAACTCTCTACGGGTAAAGCAACGGTTATACAGGAAAAATACGGTGGG 245
Db 6330  CAAAGACATTGCTCATACACGCTTAAGGCTTTAACTCAGTATCGTCGGAATGCTATTGG 6271
QY 246  CTATTTATTCAAGATTATGCCCTTGATTCGCCGACAGGACAGTTAAATTCACCTTCAGCT 305
Db 6270  TTTCGTTTTTCAATTTTATAAATTTGGTTCTTCTAATTTGACAGCTAAAGAAAAATGTTGAAT 6211
QY 306  TCGGTTGGAAAAACACAAATGGCTGAAATTCCT---CAAGTACTTTCATGCTGTGGTCT 362
Db 6210  AGCAGTTGAAATTTGTAGCAGATGCTTTAGATCCTGTGACCATTTTAAAGGAGTAGGACT 6151
QY 363  TGAGTCGTTTCGAGGAAAGCCAGTTTGTGAACTCTCTGGTGGCGAAACAACAGAACTGC 422
Db 6150  CAGTCATCGTCTGGATCAATTTCTCTGCTCAGCTCTCAGGTGGTGAACAGCAACGGGTTTC 6091
QY 423  GTTGGCCCGGTACTGCTCAAAATCCCGAATAATTCGGCTGATGAACCAACCGGAGC 482
Db 6090  GATAGCACGTGCTTACGTAAACCCCTAAATTTGCTTCTTTGTATGAACCTACAGTGC 6031
QY 483  TCTAGATTTTAAACAAACAGTGAAGTATGATAGAGCAATTGAGAGCACTCGCCGACAAAGG 542
Db 6030  CCTTGACTACCAACAGGAAGCAAAATCTTAACCTCTTACAGGATATGSCACAAACTAA 5971
QY 543  CGCCACCGTTGTTGTT 558
Db 5970  GGGGACCACCGGTAGTT 5955

```

Search completed: July 30, 2004, 13:27:45
Job time : 3638 secs

This Page Blank (uspio)

PS Claim 14; Page 23-28; 34pp; Japanese.

XX The present invention describes a protein (I) which can be used to
CC construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs
CC can be used in breeding Corynebacteria particularly for production of L-
CC glutamic acid. The present sequence encodes three ORFs (open reading
CC frames) from the Brevibacterium lactofermentum gltBD gene, which is used
CC in the exemplification of the present invention. (Updated on 15-SEP-2003
CC to standardise OS field)

SQ Sequence 2370 BP; 633 A; 563 C; 515 G; 659 T; 0 U; 0 Other;

Query Match 100.0%; Score 609; DB 3; Length 2370;
Best Local Similarity 100.0%; Pred. No. 2.2e-191;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTGGATCTTATGGCAAGGTCTC 60
DB 1117 ATGATAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTGGATCTTATGGCAAGGTCTC 1176

QY 61 AGTCATAAGTTTTACAGGAAACATGACAGCACTGACTGGAGCGTCCGGTTCAGGAAAA 120
DB 1177 AGTCATAAGTTTTACAGGAAACATGACAGCACTGACTGGAGCGTCCGGTTCAGGAAAA 1236

QY 121 TCGACTTTGCTCACTGCTTGGCACACTTGACAAACCAAGTTCGGACAGATCCCTTGTC 180
DB 1237 TCGACTTTGCTCACTGCTTGGCACACTTGACAAACCAAGTTCGGACAGATCCCTTGTC 1296

QY 181 GAGGATGAGACCTTCTGAAACTCTCTACGCGTAAGCAACGGTTATACAGGAAAAATACG 240
DB 1297 GAGGATGAGACCTTCTGAAACTCTCTACGCGTAAGCAACGGTTATACAGGAAAAATACG 1356

QY 241 GTGGCTATTTATTCAGATATGCGCTTGNATCCGACAGGACAGTTAAATTCACCTT 300
DB 1357 GTGGCTATTTATTCAGATATGCGCTTGNATCCGACAGGACAGTTAAATTCACCTT 1416

QY 301 CAGCTTGGCGTGGAAAAACAACAAATGGCTGAAATCTCTCAAGTACTTCATGCTGTGGT 360
DB 1417 CAGCTTGGCGTGGAAAAACAACAAATGGCTGAAATCTCTCAAGTACTTCATGCTGTGGT 1476

QY 361 CTGAGTCGTTGAGGAAAGCCAGATTTTGAATCTCTCTGGTGGCGCAACCAACGAACT 420
DB 1477 CTGAGTCGTTGAGGAAAGCCAGATTTTGAATCTCTCTGGTGGCGCAACCAACGAACT 1536

QY 421 CGCTTGGCCCGGTACTGCTCAAAATCCCGCAATAATTTCTGGCTGATGAACCAACCGGA 480
DB 1537 CGCTTGGCCCGGTACTGCTCAAAATCCCGCAATAATTTCTGGCTGATGAACCAACCGGA 1596

QY 481 GCTCTAGATTTAACAACAGTGAAGTATGATAGAGCATTTGAGCACTTCGCGGACAAA 540
DB 1597 GCTCTAGATTTAACAACAGTGAAGTATGATAGAGCATTTGAGCACTTCGCGGACAAA 1656

QY 541 GGCGCCACCGTGTGTGTGCTACCACTCGCCCTCTTCGAGAAATCAGCGGATACCAAT 600
DB 1657 GGCGCCACCGTGTGTGTGCTACCACTCGCCCTCTTCGAGAAATCAGCGGATACCAAT 1716

QY 601 ATCAAACTA 609

DB 1717 ATCAAACTA 1725

RESULT 2

ABQ67196_3
Continuation (4 of 7) of ABQ67196 from base 300001 (Listeria innocua contig DNA sequence
WP Sequence split into 7 fragments LOCUS ABQ67196 Accession Abq67196

WP	Fragment Name	Begin	End
WP	ABQ67196_1	1	110000
WP	ABQ67196_2	100001	210000
WP	ABQ67196_3	200001	310000
WP	ABQ67196_4	300001	410000
WP	ABQ67196_5	400001	510000
WP	ABQ67196_6	500001	610000
WP	ABQ67196_7	600001	694707

Query Match 13.3%; Score 80.8; DB 6; Length 110000;

Best Local Similarity 47.2%; Pred. No. 8.3e-15;
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTGGATCTTATGGCAAGGTCTC 60

DB 46005 ATGATAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTGGATCTTATGGCAAGGTCTC 46064

QY 61 AGTCATAAGTTTTTACCAGGAACATGACAGCACTGACTGGAGCGTCCGGTTCAGGAAAA 120

DB 46065 AGTCATAAGTTTTTACCAGGAACATGACAGCACTGACTGGAGCGTCCGGTTCAGGAAAA 46124

QY 121 TCGACTTTGCTCACTGCTTGGCACACTTGACAAACCAAGTTCGGACAGATCCCTTGTC 180

DB 46125 TCGACTTTGCTCACTGCTTGGCACACTTGACAAACCAAGTTCGGACAGATCCCTTGTC 46184

QY 181 GAGGATGAGACCTTCTGAAACTCTCTACGCGTAAGCAACGGTTATACAGGAAAAATACG 240

DB 46185 GAGGATGAGACCTTCTGAAACTCTCTACGCGTAAGCAACGGTTATACAGGAAAAATACG 46238

QY 241 GTGGCTATTTATTCAGATATGCGCTTGNATCCGACAGGACAGTTAAATTCACCTT 300

DB 46239 GTGGCTATTTATTCAGATATGCGCTTGNATCCGACAGGACAGTTAAATTCACCTT 46298

QY 301 CAGCTTGGCGTGGAAAAACAACAAATGGCTGAAATCTCTCAAGTACTTCATGCTGTGGT 360

DB 46299 CAGCTTGGCGTGGAAAAACAACAAATGGCTGAAATCTCTCAAGTACTTCATGCTGTGGT 46358

QY 361 CTGAGTCGTTGAGGAAAGCCAGATTTTGAATCTCTCTGGTGGCGCAACCAACGAACT 420

DB 46359 CTGAGTCGTTGAGGAAAGCCAGATTTTGAATCTCTCTGGTGGCGCAACCAACGAACT 46418

QY 421 GGCTTGGCCCGGTACTGCTCAAAATCCCGCAATAATTTCTGGCTGATGAACCAACCGGA 480

DB 46419 GGCTTGGCCCGGTACTGCTCAAAATCCCGCAATAATTTCTGGCTGATGAACCAACCGGA 46478

QY 481 GCTCTAGATTTAACAACAGTGAAGTATGATAGAGCATTTGAGCACTTCGCGGACAAA 540

DB 46479 GCTCTAGATTTAACAACAGTGAAGTATGATAGAGCATTTGAGCACTTCGCGGACAAA 46538

QY 541 GGCGCCACCGTGTGTGTGCTACCACTCGCCCTCTTCGAGAAATCAGCGGATACCAAT 600

DB 46539 GGCGCCACCGTGTGTGTGCTACCACTCGCCCTCTTCGAGAAATCAGCGGATACCAAT 46598

QY 601 AT 602

DB 46599 AT 46600

RESULT 3

ABQ69245_29/c
Continuation (30 of 31) of ABQ69245 from base 2900001 (Listeria innocua DNA sequence #68;
WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245

WP	Fragment Name	Begin	End
WP	ABQ69245_0	1	110000
WP	ABQ69245_01	100001	210000
WP	ABQ69245_02	200001	310000
WP	ABQ69245_03	300001	410000
WP	ABQ69245_04	400001	510000
WP	ABQ69245_05	500001	610000
WP	ABQ69245_06	600001	710000
WP	ABQ69245_07	700001	810000
WP	ABQ69245_08	800001	910000
WP	ABQ69245_09	900001	1010000
WP	ABQ69245_10	1000001	1110000
WP	ABQ69245_11	1100001	1210000
WP	ABQ69245_12	1200001	1310000
WP	ABQ69245_13	1300001	1410000
WP	ABQ69245_14	1400001	1510000
WP	ABQ69245_15	1500001	1610000
WP	ABQ69245_16	1600001	1710000
WP	ABQ69245_17	1700001	1810000

```
WP ABQ69245_18 1800001 1910000
WP ABQ69245_19 1900001 2010000
WP ABQ69245_20 2000001 2110000
WP ABQ69245_21 2100001 2210000
WP ABQ69245_22 2200001 2310000
WP ABQ69245_23 2300001 2410000
WP ABQ69245_24 2400001 2510000
WP ABQ69245_25 2500001 2610000
WP ABQ69245_26 2600001 2710000
WP ABQ69245_27 2700001 2810000
WP ABQ69245_28 2800001 2910000
WP ABQ69245_29 2900001 3010000
WP ABQ69245_30 3000001 3011208

Query Match 13.3%; Score 80.8; DB 6; Length 110000;
Best Local Similarity 47.2%; Pred. No. 8.3e-15;
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTGGCGTTCGGATCTTATGGCAAGTCTC 60
Db 34809 ATGATTAATTAATGATTAATGCTAGTAAAGATACAGATAAATGATTTGGAAAAGTC 34750
QY 61 AGTCATAAGTTTTTACCAGAACATGACAGACTGACTGGAGCGTCCGGTTTCAGAAAA 120
Db 34749 TCTCTTTCTATTGGAGCTGCGCAATTTATTGAGTCTGTCGGCAGAGTGTAGTGGGAG 34690
QY 121 TCGACTTTGCTCAACTGCTTGGCACACTTGACAAACCAAGTTCCGGACAGATCCTTGTG 180
Db 34689 ACACCTGCTTAATGTTATCGGACACCTAGATTCAAAGATAGTGGCAAGTTATTATT 34630
QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACCGGTAAGCAACGGTTTACAGSAAAAATACG 240
Db 34629 AACGAGATGGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAAAAAAGAGGTG 34576
QY 241 GTGGGCTATTATTTCAGANTTATGCTTGAATCCCGACAGCAGAGTTAAATTCACCTT 300
Db 34575 TTAGGTTTATATTCCAAAATTTATCTATTGATGGAATGAACAGTGTGAAAACCTTA 34516
QY 301 CAGCTTGGCGTGAAAAACAAATCGCCTGAAATTCCTCAAGTACTTTCATGCTGTGTGT 360
Db 34515 TCCATTACAGGTGGGGAATTCGACAGCTGATGATAGCAATTTGGAAGATGAGAAATG 34456
QY 361 CTGAGTCTGTCGAGAAAAGCAGTTTTGAACTCTCTGGTGGGGAACAAACGAACCT 420
Db 34455 GATGAGAGCTATTTTAGCAAAAAAGTATACCAATTAAGTGGTGGGAAAAACACGGATT 34396
QY 421 GCGTTGCGCGGCTACTGCTCAAAATCCCGAATTAATCTGCTGATCAACCAACCGGA 480
Db 34395 GCTATTGTGGCAATTTTACTCAAAACCAATTTCACTTTTACTTGGCAGCAACCACTGGC 34336
QY 481 GCTCTAGATTAAACAAACAGTGAAGTGTAGTCATAGAGCAATGAGACACTCGCCGACAAA 540
Db 34335 AATTAGATGATAAAACAAACAAAAATCAITGAAATTTATTTCTAGCCCTTGA AAAAGCAA 34276
QY 541 GCGGCCACCGTTGTTGTTGCTACGCACTCGCCCTCTCCGAGAAATCAGGGATACCAATT 600
Db 34275 GGTAAAGACTATCGTTTGTGTGTCAGCATGACCGGAAATATCAGGAAAAGCAGATCGGTC 34216
QY 601 AT 602
Db 34215 AT 34214
```

```
RESULT 4
ABA90521_11
Continuation (12 of 24) of ABA90521 from base 1100001 (Genomic sequence of Lactococcus
WP Sequence split into 24 fragments LOCUS ABA90521 Accession ABA90521
WP Fragment Name Begin End
WP ABA90521_00 1 110000
WP ABA90521_01 100001 210000
WP ABA90521_02 200001 310000
WP ABA90521_03 300001 410000
WP ABA90521_04 400001 510000
```

```
WP ABA90521_05 500001 610000
WP ABA90521_06 600001 710000
WP ABA90521_07 700001 810000
WP ABA90521_08 800001 910000
WP ABA90521_09 900001 1010000
WP ABA90521_10 1000001 1110000
WP ABA90521_11 1100001 1210000
WP ABA90521_12 1200001 1310000
WP ABA90521_13 1300001 1410000
WP ABA90521_14 1400001 1510000
WP ABA90521_15 1500001 1610000
WP ABA90521_16 1600001 1710000
WP ABA90521_17 1700001 1810000
WP ABA90521_18 1800001 1910000
WP ABA90521_19 1900001 2010000
WP ABA90521_20 2000001 2110000
WP ABA90521_21 2100001 2210000
WP ABA90521_22 2200001 2310000
WP ABA90521_23 2300001 2365589

Query Match 13.2%; Score 80.6; DB 6; Length 110000;
Best Local Similarity 48.9%; Pred. No. 9.6e-15;
Matches 246; Conservative 0; Mismatches 254; Indels 3; Gaps 1;

QY 59 TCAGTCATAAGTTTTTACCAGAACAAATGACAGCACTGACTGGAGCGTCCGGTTTCAGAA 118
Db 10785 TAAATTTTGAATTCAAAAGAGGAATGGCTGTCTCATTTTAGGTCTCTCTGGGGCAGGTA 10844
QY 119 ANTGCRCTTTCTCAACTGCTTGGCACACTTGACAAACCAAGTTCGGGACAGATCCTTG 178
Db 10845 AATCAACAGCTCTAAATATTTTAGAGGAATGGAATCAATATGAGGGAGAGGTTATTA 10904
QY 179 TCGAGGATGTAGACTCTGAAACTCTACGGCTAAGCAACCGGTTATACAGGAAAAATA 238
Db 10905 TTGATGAGGAAAAAATCTCTAATTTTCAAACAAAGAAATGATTACCTACCGAGTTATG 10964
QY 239 CGGTGGCGTATTATTTTCAAGATTTATGCTTGTATCCGACAGACAGATTAATTTCAACC 298
Db 10965 CCGTTGGTTTGGTTTTTCAATTTTATAATCTGTCAATAATTTGACAGCACTAGANAATG 11024
QY 299 TTCAGCTTGGGTGAAAAACAAATGGCTGAAATTCCTC---AAGTACTTTCATGCTG 355
Db 11025 TCGAATTAGCTTCTGAAATTTGTCGAAATGCTCTGTGATGCTAAAGAAGTCTGAAGATG 11084
QY 356 TTGTCCTTGAGTCTGTCGAGAAAAGCCAGTTTTTGAACCTCTCTGTGGGGGGAACAACAA 415
Db 11085 TCGGCTCTGGAACATCGGCTTCATAATTTTCTCTGCAACTTCTGTGTGGGAGCAGCAAA 11144
QY 416 GAACTGCTTGGCCCGGCTACTGCTCAAAAATCCCGAATAATTTCTGCTGATGAACCAA 475
Db 11145 GAGTAGCGATTGCGCGTGCCATGCCAAAATCCAAAATCTTGTCTTTGTGATGAACCAA 11204
QY 476 CCGGAGCTCTAGATTTAAACAAACAGTGAAGTGTATGATAGAGCAATGAGAGCACTGCGCG 535
Db 11205 CCGTGTGCTTGGAATTATCATACAGGAAAACAATCTTGAAAATCTTCAAGATATGSCCA 11264
QY 536 ACAAGGCGCCACCGTTGTGTT 558
Db 11265 GAAAAGAGGAAAAACAGTTATT 11287
```

```
RESULT 5
AAV65222/c
ID AAV65222 standard; DNA; 1010 BP.
XX
XX AAV65222;
AC
XX AAV65222;
DT 24-DEC-1998 (first entry)
DX
XX DNA encoding a S. pneumoniae transport protein.
DE
XX Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
KW virulence; antibody; infection; detection; treatment; ss.
```

XX OS Streptococcus pneumoniae.
 XX FN WO9826072-A1.
 XX PD 18-JUN-1998.
 XX PF 09-DEC-1997; 97WO-US022578.
 XX PR 13-DEC-1996; 96US-0036281P.
 XX PA (ELIL) LILLY & CO ELI.
 XX PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;
 XX PI Mills BJ, Norris FH, Peery RB, Rostek PR, Skatrud PL;
 XX PI Smith MC, Solenberg PJ, Treadway PJ, Young Bellido ML;
 XX DR WPI: 1998-348529/30.
 XX DR P-PSDB; AAW80648.
 XX ST Streptococcus pneumoniae nucleic acid sequences - used in DNA chips for
 PT evaluating gene expression, and identification of virulence genes.
 XX PF
 XX PS Claim 1; Page 63; 333pp; English.
 XX CC This DNA sequence encodes a Streptococcus pneumoniae transport protein.
 CC The invention provides DNA sequences (AAV65201 to AAV65304) from the
 CC Streptococcus pneumoniae genome and corresponding protein sequences
 CC (AAW80605 to AAW80728). A recombinant host containing a vector comprising
 CC any of the above nucleic acids can be used for the recombinant expression
 CC of the protein sequences. The invention also provides a DNA chip having
 CC arrayed on it at least 15 base pair fragment of any one or more of these
 CC DNA sequences. The DNA chip can be used for evaluating gene
 CC expression in S. pneumoniae and for identifying virulence genes in S.
 CC pneumoniae. Antibodies that selectively bind to the above proteins or
 CC peptide fragments can be used to treat S. pneumoniae infection. The
 CC antibodies can also be used to detect S. pneumoniae cells
 XX SQ
 SQ Sequence 1010 BP; 285 A; 247 C; 207 G; 271 T; 0 U; 0 Other;
 Query March 13.2; Score 80.4; DB 2; Length 1010;
 Best Local Similarity 49.8; Pred. No. 9.9e-16;
 Matches 231; Conservative 0; Mismatches 231; Indels 2; Gaps 1;
 QY 55 GGTCTCAGTCATAGTTTATACAGGAAACAAATGACAGCACTGACGAGCGTCCGGTTCA 114
 DB 1007 GATGTGAATTTGAGATTGAAGGGGGAAGTGGTTATTATCTTGGTCTTCAGGTGCA 948
 QY 115 GGAAATCGAGTTTGTCTCACTGCTTGGCACACTTGCAAAACAGATTCGGGACAGATC 174
 DB 947 GCGAAGTCAACAGTCTTAACCTTCTTGGGGGAATGATACCAATGATGAAGGGGAATC 888
 QY 175 CTTCTCAGGATGATGAGACCTCTGAAACTCTCTACGGGTAAAGCAACGGTTATACAGCAA 234
 DB 887 TGGATTGATGGTGTATATATGGGATATAGTTCACAGCGCAATACCTAGCA 828
 QY 235 AATACGTTGGCTATTTATTTCAAGATATGCTTGAATCCGACAGGACAGTTAAATTC 294
 DB 827 AATGATGTGGGTTTGTCTTCACTGTTTAAATCTAGTTTCTAATCTGACAGTAAGGAA 768
 QY 295 AACCTTCAGTTCGGGGGAAACAAATGCCTGAAATTC-CTCAAGTACTTCTAG 352
 DB 767 AATGTGAATGCTTCTGAATTTGTGACATGCTTGAATTTGATCAGGCTTTCAGAG 708
 QY 353 CTGTGCTTGTAGTCTGTTCCAGAAAGACCGAGTTTGTGAATCTCTGTTGGCGAACAAC 412
 DB 707 ATGTAGTCTGCTCATCGTCTAAATTAACCTTCCAGCCCAAGCTTCTGAGGGGAGCAAC 648
 QY 413 AACGAATCGTGGCCCGGCTACTGCTCAAAATCCCGAATATCTGGCTGATGAAC 472
 DB 647 AGCGAGTCTCATTTGACGCGCGGTAGCCAAATCTTAAATTTCTCTTTGTGTGAAC 588
 QY 473 CAACCGGAGCTCTAGATTTTAAACAAAGTGAAGTCTAGTATAGAA 516

DB 587 CGACTGGAGCCTTGGATTATCAGACGGCAAGCAGGTTTGGAAA 544
 RESULT 6
 ID AEN66720 standard; DNA; 711 BP.
 XX AC AEN66720;
 XX DT 01-JUL-2002 (first entry)
 XX DE Streptococcus polynucleotide SEQ ID NO 1353.
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX OS Streptococcus pyogenes.
 XX PN WO200234771-A2.
 XX PD 02-MAY-2002.
 XX PF 29-OCT-2001; 2001WO-GB004789.
 XX PR 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
 PI Tettelin H;
 XX WPI; 2002-352536/38.
 DR P-PSDB; ABP26089.
 XX New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX Claim 7; Page 3291; 4525pp; English.
 XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), AEN66044-ABN71546 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX SQ Sequence 711 BP; 211 A; 130 C; 159 G; 211 T; 0 U; 0 Other;
 Query Match 13.1; Score 80; DB 6; Length 711;
 Best Local Similarity 49.0; Pred. No. 1.1e-15;
 Matches 243; Conservative 0; Mismatches 250; Indels 3; Gaps 1;
 QY 66 TAAGTTTTTACAGGAACTGACAGCACTGACGAGCGTCCGGTTCAGGAAATCGAC 125
 DB 96 TGAGATTAAAGAGAACTAGTGGTATTTTGGGAGCATCCGAGCTGGTAAATCGAC 155

QY	126	TTTGCTCACTGCTTGGGACACTTGACAAACCAAGTTCGGGACAGATCCTTGTGAGGA	185
Db	156	ATGCTCTCAATATTTTGGGGGCAAGGACACAGTAGAGGTCAGTAAATTAATGATGG	215
QY	186	TGTAGACCTTCTGAAACTCTCTACGCTAAGCAACGTTTATACAGGAAATAACGTTGGG	245
Db	216	CAAGAACAATTGCTCAATACACGCTTAAGGCTTAACTCAGTATCGTTCGGAATGCTATTGG	275
QY	246	CTATTATTATTCRAGATATGCTTGTGATTCCTCGACAGGACAGTTAAATTTCAACCTTCAGCT	305
Db	276	TTTTCGTTTTTCAATTTTATTAATTTGGTCTCTAAATTTGACAGCTTAAAGAAATGTTGAAT	335
QY	306	TGCGGTGGAAAAACACAAATGGCTGAAATTCCT---CAAGTACTTCACTGCTGTTGGTCT	362
Db	336	AGCAGTTGAAATTTGACAGATGCTTTAGATCTGTGACCAATTTTAAAGGAAGTAGGACT	395
QY	363	TGAGTCGTTCCAGAAAAAGCCAGTTTTTGAACCTCTGTTGGGGAACACACAGCACTGC	422
Db	396	CAGTCATCGCTCGATCATTTTCTCTCTCAGCTCTCAGGTGGTGAACGCAACGGGTTTC	455
QY	423	GTTCGGCCCGGTACTGCTCAAAATCCCCGAATAATTCCTGGCTGATGAACCAACCGGAGC	482
Db	456	GATAGCAGTGCCTTAGCTAATAAACCCATAATTTGCTTCTTTGTGATGAACCTACAGGTGC	515
QY	483	TCTAGATTTAAACACAGTAGCTAGTCTATAGAACGATTTAGAGCACTCGCGCAAAAGG	542
Db	516	CCTTGACTACCAACAGGAAGCAAAATCTTAACCTCTTACAGGATATGGCACAACATAA	575
QY	543	CGCCACCGTTGTGTT 558	
Db	576	GGGACACAGGTAGTT 591	

RESULT 7
ABA03041_28/c
Continuation (29 of 30) of ABA03041 from base 2800001 (Listeria monocytogenes BGD-e Gene
WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041

WP	Fragment Name	Begin	End
WP	ABA03041_00	1	110000
WP	ABA03041_01	100001	210000
WP	ABA03041_02	200001	310000
WP	ABA03041_03	300001	410000
WP	ABA03041_04	400001	510000
WP	ABA03041_05	500001	610000
WP	ABA03041_06	600001	710000
WP	ABA03041_07	700001	810000
WP	ABA03041_08	800001	910000
WP	ABA03041_09	900001	1010000
WP	ABA03041_10	1000001	1110000
WP	ABA03041_11	1100001	1210000
WP	ABA03041_12	1200001	1310000
WP	ABA03041_13	1300001	1410000
WP	ABA03041_14	1400001	1510000
WP	ABA03041_15	1500001	1610000
WP	ABA03041_16	1600001	1710000
WP	ABA03041_17	1700001	1810000
WP	ABA03041_18	1800001	1910000
WP	ABA03041_19	1900001	2010000
WP	ABA03041_20	2000001	2110000
WP	ABA03041_21	2100001	2210000
WP	ABA03041_22	2200001	2310000
WP	ABA03041_23	2300001	2410000
WP	ABA03041_24	2400001	2510000
WP	ABA03041_25	2500001	2610000
WP	ABA03041_26	2600001	2710000
WP	ABA03041_27	2700001	2810000
WP	ABA03041_28	2800001	2910000
WP	ABA03041_29	2900001	294528

Query Match 13.0%; Score 79.2; DB 6; Length 110000;
Best Local Similarity 47.0%; Pred. No. 2.8e-14;
Matches 283; Conservative 0; Mismatches 313; Indels 6; Gaps 1;

QY	1	ATGATAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTCGATCTTTATGCGAAGGTCTC	60
Db	58203	ATGATTTGAAATTTAGTTTAAATTTAGTTCAGTTAAAGATACAGATAAATTTGTTGAAAAAGTC	58144
QY	61	AGTCATAGTTTTCACAGGAACAATGACAGCACTGAGTGGAGCGTCCGGTTCAGGAAAA	120
Db	58143	TCTCTTTTATTTGGAGCTGGTGAATTTATTGCACTCGTTCGGCGAGAGTGGTAGTGGGAAG	58084
QY	121	TGCACTTTGCTCAACTGCTTTGGCACACTTGACAAACCAAGTTCGGGACAGATCCTTGTCT	180
Db	58083	ACCACGCTGCTAAATGTTATCGGACACCTAGATTCAAAAGATAGTGGGCAAGTTATTATT	58024
QY	181	GAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCAACCGTGTATACAGGAAAAATACG	240
Db	58023	AACGAGATGGAATATCAG-----CGAAAAAGAGGTTATGACTCTAAAAAAAAGAGGTG	57970
QY	241	GTGGGCTATTTTATTTCAAGATTATGCCCTTGATTTCCCGACAGGACAGTTAAATTCACCTT	300
Db	57969	TTAGGTTTATTTTCCAAAATTTATCTATTGATGAAAAATGAACAGTGTTTAGAAAACTTA	57910
QY	301	CAGCTTGGCGTGGAAAAACACAAATGGCTGAAATTCCTCAAGTACTTCTGCTGTTGGT	360
Db	57909	TCCATTACAGTGGGAAAAATCGCAAGCTGATGATAGAGCATTTAGAAGTAGGATG	57850
QY	361	CTTGAGTGTTCGAGGAAAAAGCCAGTTTTTGAACCTCTCTGGTGGCGAACCAACAGAACT	420
Db	57849	GATGAGAGCTATTTAGCAAAAAAAGTATACCAATTAAGCGGTGGAGAAAAACAACGGATT	57790
QY	421	CGCTTGGCCCGGTACTGCTCAAAAAATCCCGAATAATTTCTGGCTGATGAACCAACCGGA	480
Db	57789	GCTATTGTGCGCATTTTACTCAAACCATTTCACTGTTTACTTGGGACGACCAACAGGC	57730
QY	481	GCTCTAGATTTAACAAACAGTGAGTCTAGTCTATAGAAGCATTTGAGACACTGCGCGACAAA	540
Db	57729	AACTTGGATGATAAAAAACAAACAAAAATCATTTGAATTTATTTTAGCTTTGAAAAAGCAA	57670
QY	541	GGCGCCACCGTTGTTGTTGCTACGCACTCGCCCTCTTCCGAGATCAGCGGATACCAT	600
Db	57669	GGTAAGACTATCGTTTGTGTTTACGATGACCCAGAAATATCTGCAAAAGCAGATCGAATC	57610
QY	601	AT 602	
Db	57609	AT 57608	
RESULT 8			
AAS55747			
ID	AAS55747 standard; DNA; 711 BP.		
XX	AAS55747;		
AC	AAS55747;		
XX	AAS55747;		
DT	13-FEB-2002 (first entry)		
XX	Streptococcus pneumoniae DNA for cellular proliferation protein #318.		
DE	Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;		
XX	antibacterial; drug design.		
XX	Streptococcus pneumoniae.		
OS	WO200170955-A2.		
PN	27-SEP-2001.		
XX	21-MAR-2001; 2001WO-US009180.		
PF	21-MAR-2000; 2000US-0191078P.		
XX	23-MAY-2000; 2000US-0206848P.		
PR	26-MAY-2000; 2000US-0207727P.		
PR	23-OCT-2000; 2000US-0242578P.		
PR	27-NOV-2000; 2000US-0253625P.		
PR	22-DEC-2000; 2000US-0257931P.		
PR	16-FEB-2001; 2001US-0269308P.		

XX PA (ELIT-) ELITPA PHARM INC.
 XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
 XX PI Yamamoto RT, Xu HH;
 XX DR WPI: 2001-611495/70.
 XX DR P-PSDB; AAU37888.
 XX PT New polynucleotides for the identification and development of
 XX PT antibiotics, comprise sequences of antisense nucleic acids.
 XX XX Claim 27; SEQ ID NO 9384; Slipp; English.
 XX XX The invention relates to antisense inhibitors of genes essential to
 XX CC prokaryotic cellular proliferation, their use in identifying the genes,
 XX CC their use in the discovery of novel antibiotics, the essential genes
 XX CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 XX CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 XX CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 XX CC useful for the identification of potential new targets for antibiotic
 XX CC development. The antisense nucleic acids can also be used to identify
 XX CC proteins used in proliferation, to express these proteins, and to obtain
 XX CC antibodies capable of binding to the expressed proteins. The proteins can
 XX CC be used to screen compounds in rational drug discovery programmes. The
 XX CC antisense nucleic acid sequence is also useful to screen for homologous
 XX CC nucleic acids which are required for cell proliferation in a wide variety
 XX CC of organisms. The present sequence encodes an essential prokaryotic
 XX CC cellular proliferation protein. Note: The sequence data for this patent
 XX CC did not form part of the printed specification, but was obtained in
 XX CC electronic format directly from WIPO at
 XX CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 711 BP; 187 A; 132 C; 190 G; 202 T; 0 U; 0 Other;
 Query Match 12.7%; Score 77.4; DB 4; Length 711;
 Best Local Similarity 50.8%; Pred. No. 8.2e-15;
 Matches 236; Conservative 0; Mismatches 226; Indels 3; Gaps 2;
 QY 55 GGCTCAGTCAATAGTTTACACAGGAACAATGACAGACTGACTGGAGCGTCCGGTTCA 114
 Db 82 GATGTGAATTTGAGATTGAAGAGGGGAACTGGTTATTATCTCTTGGTCTTCAGGTGCA 141
 QY 115 GGAAATCGACTTTGCTCACTGCTTTGGCACACTTGCACAAACCAAGTTCGGACAGATC 174
 Db 142 GCGAAGTCAACAGTCTTAACCTTCTTGGGGGAATGGTACCAATGATGAGGGGAATC 201
 QY 175 CTGTGCGAGGATGAGACCTTCTGAACACTCTCTACGGGTGAAGCAACGGTTATACAGGAA 234
 Db 202 TGGATTGATGGTGTATATTGCGGATTATAGTTCCACAGCGCAACCAATTACCGTAGA 261
 QY 235 AATACGGTGGGCTATTTATTCAGATATGCTTGTGATTCCTCCGACAGGACAGTAATTC 294
 Db 262 AATGATGTTGGGTTGTTTTCAGTTTATTAATAGTTTCTATCTGACAGTAAAGGAA 321
 QY 295 AACCTTCAGCTTGGCGTGGAAA--AACACAAATGGCCTGAAAT-TCCTCAAGTACTTCAT 351
 Db 322 AATGTGAACCTGGCTCTGAAATTTGTGACAGATGCCCTTGAATTCGATCAGGCTTGACA 381
 QY 352 GCTGTGTTCTTGTAGTGTTCGAGGAAAGCCAGTTTGTGAATCTCTGTGTCGGACAA 411
 Db 382 GATGTAGTCTGGTCTATCGTCTAAATAACTTTCCAGCCAGCTTTCTGAGGGGAGCAA 441
 QY 412 CAACGAACCTGCTGGCCCGGCTACTGCTCAAAAATCCCGAATAATTCCTGGCTGATGA 471
 Db 442 CAGCGAGTCTCCATGTGACGCGCGGTAGCCAAATCTTAATATTCCTCTTGTGATGA 501
 QY 472 CCACCGGAGCTCTAGATTTTAAACAAACAGTAGGCTAGTCTAGAA 516
 Db 502 CCGACTGGAGCTTGGATTATCAGACGGGCAAGCAGGTTTGTGAA 546

ABN67102
 ID ABN67102 standard; DNA; 750 BP.
 XX AC ABN67102;
 XX DT 01-JUL-2002 (first entry)
 XX DE Streptococcus polynucleotide SEQ ID NO 2117.
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 XX KW antinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX OS Streptococcus agalactiae.
 XX PN WO200234771-A2.
 XX PD 02-MAY-2002.
 XX PF 29-OCT-2001; 2001WO-GB004789.
 XX PR 27-OCT-2000; 2000GB-00026333.
 XX PR 24-NOV-2000; 2000GB-00028727.
 XX PR 07-MAR-2001; 2001GB-00005640.
 XX PA (CHIR-) CHIRON SPA.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
 XX PI Tettelin H;
 XX WPI: 2002-352536/38.
 XX DR P-PSDB; ABP26471.
 XX PT New Streptococcus protein for the treatment or prevention of infection or
 XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
 XX PT detecting a compound that binds to the protein.
 XX PS Claim 7; Page 3364; 4525pp; English.
 XX CC The invention relates to a protein (ABP25413-BP30895) from group B
 XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 XX CC the specification. The proteins have antibacterial and antinflammatory
 XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 XX CC antibodies that bind (I) are used in the manufacture of medicaments for
 XX CC the treatment or prevention of infection or disease caused by
 XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 XX CC biological sample. (I) is used to determine whether a compound binds to
 XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 XX CC used as a vaccine or diagnostic composition. The disease caused by
 XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
 XX CC used in gene therapy. Antibodies to (I) are used for affinity
 XX CC chromatography, immunoassays, and distinguishing/identifying
 XX CC Streptococcus proteins
 XX SQ Sequence 750 BP; 249 A; 119 C; 143 G; 239 T; 0 U; 0 Other;
 Query Match 12.4%; Score 75.6; DB 6; Length 750;
 Best Local Similarity 48.3%; Pred. No. 3.3e-14;
 Matches 259; Conservative 0; Mismatches 259; Indels 18; Gaps 1;
 QY 52 CAAGTCTCAGTCATAGTATTTTACAGGAACAATGACAGCACTGACTGGAGCTCCGGT 111
 Db 73 CAAGAGCTGATTTTAAAGTTGAACAGGGGAATTCATTGCTATTATGGGAGAGTCTGT 132
 QY 112 TCAGAAATTCAGCTTTGCTCACTGCTTGGACACTTGCACATTCGACAAACCAAGTCCGGACAG 171
 Db 133 TCGGAAAAACAACACTCTCTGCTAAATATTTTACTTCTTTAGAAAAACCGCAATGGACAA 192
 QY 172 ATCTTGTGCGAGGATGTAGACCTTCTGAACTCTCTACCGCTAAGCAACGGTTATACAGG 231

Db 193 GTGATTAAATGGGAGATATACGAAATTAAGAGGCAAAATAGGAGTTTCGT 252
Qy 232 AAAAAATACGGTGGCTATTTATTTCAAGATATATGCTTATCCCGACAGGACAGTAA 291
Db 253 TTGAAAAATCTGGTTTCGTCCTTTCAGGACTTCAAGCTTTTGATATCTCTCTGTAA 312
Qy 292 TTCAACCTTCAGCTTCGGTGG-----AAAAACACAATGGCTGAA 333
Db 313 GATTAATATCTATCTTCTTTAGTTCTTGATCGTAAAGGTTACAAAGATATGATCACT 372
Qy 334 ATTCCTCAAGTACTTCATGCTGTGTGCTGTAGTCTGTCGAGAAAGCCAGTTTGA 393
Db 373 TTGTCAGAAATATCTTCTCATCTGAGAAATGATGACTTATTAGATAAGAGACCTTTCGAG 432
Qy 394 CTCTCTGGTGGCAACAAAGAACTGCTGTGGCCGGGTACTGCTCAAAATCCCGA 453
Db 433 CTCTCTGGTGGCAACAAAGAACTGCTGTGGCCGGGTACTGCTCAAAATCCCGA 492
Qy 454 ATAAATCTGCTGATGAACCAACCGGAGCTCTAGATTTAAACAAAGTGTAGTACATA 513
Db 493 ATTTTATTACAGATGAACCAACAGAGCGTTAGATTACGTTAATTCAGAGACTTGTCTG 552
Qy 514 GAAGCATTTAGAGACTTCGCCGCAAAAGCGCCACCGTTGTTGCTTCTAGGACTC 569
Db 553 AATTTATTGAAACTATTACTTTGGATGGACAAACTATTTTTGATGTAAACCCATTC 608

RESULT 10
ABN71527_09
Continuation (10 of 22) of ABN71527 from base 900001 (Streptococcus polynucleotide SEQ 1)
WP Sequence split into 22 fragments LOCUS ABN71527 Accession ABN71527
Fragment Name Begin End
WP ABN71527_01 1 110000
WP ABN71527_02 100001 210000
WP ABN71527_03 200001 310000
WP ABN71527_04 300001 410000
WP ABN71527_05 400001 510000
WP ABN71527_06 500001 610000
WP ABN71527_07 600001 710000
WP ABN71527_08 700001 810000
WP ABN71527_09 800001 910000
WP ABN71527_10 900001 1010000
WP ABN71527_11 100001 1110000
WP ABN71527_12 1100001 1210000
WP ABN71527_13 1200001 1310000
WP ABN71527_14 1300001 1410000
WP ABN71527_15 1400001 1510000
WP ABN71527_16 1500001 1610000
WP ABN71527_17 1600001 1710000
WP ABN71527_18 1700001 1810000
WP ABN71527_19 1800001 1910000
WP ABN71527_20 1900001 2010000
WP ABN71527_21 2000001 2110000
WP ABN71527_22 2100001 2155561

Query Match 12.48; Score 75.6; DB 6; Length 110000;
Best Local Similarity 48.3%; Pred.No. 4.4e-13;
Matches 259; Conservative 0; Mismatches 259; Indels 18; Gaps 1;

Qy 52 CAAGGCTCGATCATAGTTTATACAGAGCAACATGACAGCTGATGAGCGTCCGGT 111
Db 51728 CAAGACGTCGATTTAAGTTGACAGAGGGGAATTCATTGCTATTATGGAGAGTCTGGT 51787
Qy 112 TCAGAAATTCAGTTTGTCTCACTGTCTTGGCACACTTGACAAACAGTTCGGACAG 171
Db 51788 TCGGAAAAACAACTCTGCTAAATATTATTAGTACTTTAGAAAAACCGCAATATGACAA 51847
Qy 172 ATCTCTGTCAGGATGTAGACCTTCTCAACTCTCTACGGTGAAGCAACGGTTATACAG 231
Db 51848 GTGATTTTAAATGGGAGATATTAACGAAATTAAGAGGCAAAATAGGAGTTTCGT 51907
Qy 232 AAAAATACGGTGGCTATTTATTTCAGATTTATGCTTGTATCCCGACAGGACAGTAA 291

Db 51908 TTGAAAAATCTTGGTTTGGTCTTTTCAGGACTTCAACCTTTTGGATCTCTCTGTAAGA 51967
Qy 292 TTCACCTTCAGCTTGGGTGG-----AAAAACACAATGGCCTGAA 333
Db 51968 GATAATATCTATCTCTCTTTAGTTCTTGATCGTAAAGTTCACAAAGAAATGGATCATCGT 52027
Qy 334 ATTCCTCAAGTACTTCATGCTGTGTGCTGTAGTCTGTCGAGGAAAGCCAGTTTTCGAA 393
Db 52028 TTGTCAGAAATATCTTCTCATCTGAGAAATGATGACTTATTAGATAAGAGACCTTTCGAG 52087
Qy 394 CTCTCTGGTGGCAACAAAGAACTGCTGTGGCCGGGTACTGCTCAAAATCCCGA 453
Db 52088 CTCTCTGGTGGCAACAAAGAACTGCTGTGGCCGGGTACTGCTCAAAATCCCGA 52147
Qy 454 ATAAATCTGCTGATGAACCAACCGGAGCTCTAGATTTAAACAAAGTGTAGTACATA 513
Db 52148 ATTTTATTACAGATGAACCAACAGAGCGTTAGATTACGTTAATTCAGAGACTTGTCTG 52207
Qy 514 GAAGCATTTAGAGACTTCGCCGCAAAAGCGCCACCGTTGTTGCTTCTAGGACTC 569
Db 52208 AATTTATTGAACTATTACTTTGGATGGACAAACTATTTTTGATGTAAACCCATTC 52263

RESULT 11
ABX07412
ID ABX07412 standard; DNA; 708 BP.
XX
AC ABX07412;
XX
DT 27-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain coding region #1700.
XX
KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antiinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.
XX
OS Streptococcus pneumoniae; type 4 strain.
XX
PN WO200277021-A2.
XX
PD 03-OCT-2002.
XX
PF 27-MAR-2002; 2002WO-IB002163.
XX
PR 27-MAR-2001; 2001GB-00007658.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Masignani V, Tettelin H, Fraser C;
XX
DR WPI; 2003-040579/03.
DR P-PSDB; ABU02123.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae.
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX
PS Claim 6; SEQ ID NO 3399; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB556454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a

CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to Streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2489
 CC identified coding region from the genomic sequence. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
 CC standardise OS field)

XX Sequence 708 BP; 184 A; 134 C; 190 G; 200 T; 0 U; 0 Other;

Query Match 12.4%; Score 75.4; DB 7; Length 708;
 Best Local Similarity 50.4%; Pred. No. 3.8e-14;
 Matches 212; Conservative 0; Mismatches 206; Indels 3; Gaps 1;
 QY 99 TGGAGCGTCCGGTTCAGGAAATCGACTTTGCTCAACTGTCTGGCACACTTGACAAACC 158
 Db 126 TGGTCTTCAGGTGAGGCAAGTCAACAGTCTTAACTCTTGGGGGAATGATACCAA 185
 QY 159 AAGTCCGACAGATCTTGTGAGGATGTAGACTTCTGAACTCTCTACGCGTAGCA 218
 Db 186 TGATGAAGGGGAATCTGGATTGATGTGTTAATTTGGCGATTATAGTCCACACGCG 245
 QY 219 ACGGTTATACAGGAAATACGGTGGGCTATTTATTTCAAGATTATGCTTGAATCCCGA 278
 Db 246 CACCAATACCGTAGAATGATGGGGTTGTTTTCAGTTTATATCTAGTTCTAA 305
 QY 279 CAGGACAGTTAAATCAACTTACGTTGGGTGGGAAACACAAATGGCTGAATTC 338
 Db 306 TCTGACAGCTAAGGAAATGGAACGTGCTCTGAAATTTGTGACAGATCCCTTGAATCC 365
 QY 339 ---TCAAGTACTTCATGCTGTGTGTTGCTTGTAGTGTGTCGAGAAAGCCAGTTTGTGA 395
 Db 366 TGATCAGCCCTTGACAGATGATGCTTGCTCATCTCTCAATTAATTTCCAGCCAGCT 425
 QY 396 CTCTGTGGCGAACAACACGAACTGGCTTGGCCCGGTACTGCTCAAAATCCCGAAT 455
 Db 426 TCTTGGAGGGAGCAACACGAGCTCTCAATGACGCGCGGTAGCCAAAATCTCAAAAT 485
 QY 456 AATTCTGCTGATGACCAACCGGAGCTCTAGATTTTACAAACAGTACGCTAGCTAGA 515
 Db 486 TCTCCTTTGTGATGAACCGACTGAGGCTTGGATTATACAGACGGGCAAGCAGGTTTGA 545
 QY 516 A 516
 Db 546 A 546

RESULT 12

AAV52169
 ID AAV52169 standard; DNA; 21706 BP.

XX AC AAV52169;

XX DT 23-OCT-1998 (first entry)

XX DE Streptococcus pneumoniae genome fragment SEQ ID NO:36.

XX KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.
 OS WO9818931-A2.
 PN 07-MAY-1998.
 PD 30-OCT-1997; 97WO-US019588.
 PF 31-OCT-1996; 96US-0029960P.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA Kunsch CA, Choi GH, Dillon P, Rosen CA, Barash SC, Fannon M;
 PI Dougherty BA;
 XX WPI; 1998-272225/24.
 DR Computer-readable medium with recorded Streptococcus pneumoniae
 XX polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
 XX Claim 1; Page 358-371; 1409pp; English.
 CC The present invention describes a computer readable medium which has the
 CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
 CC on it, or a representative fragment or a sequence at least 85% identical
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
 CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
 CC pneumoniae. The present invention also describes an isolated nucleic acid
 CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
 CC by a process comprising: (a) screening a genomic DNA library using as a
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
 CC 391, identifying members of the library which contain sequences that
 CC hybridise to the target sequence and isolating the nucleic acid molecules
 CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an
 CC organism, amplifying nucleic acid molecules whose nucleotide sequence is
 CC homologous to amplification primers derived from the fragment of the S.
 CC pneumoniae genome to prime the amplification and isolating the amplified
 CC sequences. The computer readable medium can be used in a computer-based
 CC system for identifying fragments of the S. pneumoniae genome of
 CC commercial importance, or expression modulating fragments of the S.
 CC pneumoniae genome. Products from the present invention can be used in
 CC diagnosis kits and assays, and pharmaceutical compositions and vaccines
 CC for S. pneumoniae

XX Sequence 21706 BP; 6489 A; 3946 C; 4879 G; 6392 T; 0 U; 0 Other;

Query Match 12.4%; Score 75.4; DB 2; Length 21706;
 Best Local Similarity 50.4%; Pred. No. 2.2e-13;
 Matches 212; Conservative 0; Mismatches 206; Indels 3; Gaps 1;
 QY 99 TGGAGCGTCCGGTTCAGGAAATCGACTTTGCTCAACTGTCTGGCACACTTGACAAACC 158
 Db 14961 TGGTCTTCAAGTGTGAGGCAAGTCAACAGTCTTAACTCTTGGGGGAATGGAATACCA 15020
 QY 159 AAGTTCGCGACAGATCCCTGTGAGGATGTAGACCTTCTGAAACTCTCTACGCGTAGCA 218
 Db 15021 TGATCAAGGGGAATCTGGATTGATGGTGTAAATTTGCGGATTATAGTTCGCCACGCG 15080
 QY 219 ACGGTTATACAGGAAATACGTTGGGCTATTTATTTCAAGATTATGCTTATCCGA 278
 Db 15081 CACCAATACCGTAGAATGATGTGGGTTTGTTCAGTTTATATCTAGTTCTAA 15140
 QY 279 CAGGACAGTTAAATTCACACTTCAGCTTCCGGTGGGAAACACAAATGGCTGAATTC 338
 Db 15141 TCTGACAGCTAAGGAAATGTGGAACCTGCTTCTGAAATTTGTGACAGATGCTTGAATTC 15200
 QY 339 ---TCAAGTACTTCACTGCTGTGTGTTGCTTGTAGTCTGTTGAGGAAAGCCAGTTTGTGA 395
 Db 15201 TGATCAGCCCTTGACAGATGATGAGTCTGGCTCATCTGCTCAATAACTTTCCAGCCAGCT 15260

QY 396 CTCTGTGCGGCAACAAACGAACTGGCTGGTGGCCGGGTACTGCTCAAAATCCCGAAT 455
 Db 15261 TTCTGGAGGGAGCAACAGCGAGTCTCCATTGACGCGGGTAGCCAAAATCTTANAAT 15320
 QY 456 AATTCCTGGCTGATGACCAACCGAGCTCTAGATTAAACAAACAGTGTAGTCATAGA 515
 Db 15321 TCTCCTTTGTGATGAACCGACTGGAGCCCTTGATATATCAGACGGGCAAGCAGGTTTGA 15380
 QY 516 A 516
 Db 15381 A 15381

RESULT 13
 ABS56454_14/c
 Continuation (15 of 22) of ABS56454 from base 1400001 (Streptococcus pneumoniae type 4
 WP Sequence split into 22 fragments LOCUS ABS56454 Accession Abs56454
 Fragment Name Begin End
 WP ABS56454_00 1 110000
 WP ABS56454_01 100001 210000
 WP ABS56454_02 200001 310000
 WP ABS56454_03 300001 410000
 WP ABS56454_04 400001 510000
 WP ABS56454_05 500001 610000
 WP ABS56454_06 600001 710000
 WP ABS56454_07 700001 810000
 WP ABS56454_08 800001 910000
 WP ABS56454_09 900001 1010000
 WP ABS56454_10 1000001 1110000
 WP ABS56454_11 1100001 1210000
 WP ABS56454_12 1200001 1310000
 WP ABS56454_13 1300001 1410000
 WP ABS56454_14 1400001 1510000
 WP ABS56454_15 1500001 1610000
 WP ABS56454_16 1600001 1710000
 WP ABS56454_17 1700001 1810000
 WP ABS56454_18 1800001 1910000
 WP ABS56454_19 1900001 2010000
 WP ABS56454_20 2000001 2110000
 WP ABS56454_21 2100001 2162598

Query Match 12.4%; Score 75.4; DB 7; Length 110000;
 Best Local Similarity 50.4%; Pred. No. 5.2e-13;
 Matches 219; Conservative 0; Mismatches 206; Indels 3; Gaps 1;
 QY 99 TGGAGGCTCGGTTTCAGGAAATCGACTTGTGCTCACTGCTTGGCAGACTTCACAAAC 158
 Db 48073 TGGTCTTCAGTGCAGGCAAGTCAACAGTCTTACCTTCTTGGGGATGATACCAA 48014
 QY 159 AAGTTCGGACAGATCCTTTCGAGGATGAGACCTTCTCAAACTCTCTACGGGTAGCA 218
 Db 48013 TGATGAAGGGGAATCTGGATTGATGTTGTTAATTTGCGGATTATAGTTCACACGCG 47954
 QY 219 ACGTTTATACAGGAATAACGGTGGCTATTTATTTCAAGATTATGCTTGTATCCCGA 278
 Db 47953 CACCAATTACGTTAGAAATGATGGGGTTTGTCTTTCAGTTTATATCTAGTTCTAA 47894
 QY 279 CAGGACAGTTAAATTCACCTTCAGCTTGGGTTGGGAAAAACAAATGGCTGAATTC 338
 Db 47893 TCTGACAGTAAAGAAATGGAACCTGCTTCTGAAATTTGTGACAGATGCTTGAATCC 47834
 QY 339 ---TCAAGTACTTCATGCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 395
 Db 47833 TGATCAGGCTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 47774
 QY 396 CTCTGGTGGCAACAAACGAACTGGCTGGTGGCCGGGTACTGCTCAAAATCCCGAAT 455
 Db 47773 TTCTGGAGGGAGCAACAGCGAGTCTCCATTGACGCGGGTAGCCAAAATCTTANAAT 47714
 QY 456 AATTCCTGGCTGATGACCAACCGAGCTCTAGATTAAACAAACAGTGTAGTCATAGA 515
 Db 47713 TCTCCTTTGTGATGAACCGACTGGAGCCCTTGATATATCAGACGGGCAAGCAGGTTTGA 47654

QY 516 A 516
 Db 47653 A 47653
 RESULT 14
 ABQ71073/c
 ID ABQ71073 standard; DNA; 12278 BP.
 XX AC ABQ71073;
 XX 29-AUG-2003 (revised)
 DT 29-AUG-2002 (first entry)
 XX Listeria monocytogenes 4b contig DNA sequence #1015.
 XX Antibacterial; Listeria; food contamination; mutational analysis;
 KW infection; ds.
 XX Listeria monocytogenes ATCC 19115.
 XX WO200228891-A2.
 XX PD 11-APR-2002.
 XX 04-OCT-2001; 2001WO-FR003061.
 PF 04-OCT-2000; 2000FR-00012697.
 PR (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX Kunst F, Glaser P;
 XX WPI; 2002-332479/37.
 DR New genomic sequences from Listeria species, useful for detection,
 XX treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators.
 XX Claim 14; SEQ ID NO 3886; 180pp; French.
 XX The present invention relates to nucleic acid sequences (ABQ67188-
 CC ABQ71212) from Listeria sp. The sequences are useful as probes and
 CC primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of gene
 CC expression. Proteins encoded by the nucleic acid sequences can be used to
 CC screen for compounds that modulate gene expression, replication and
 CC pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in anti-
 CC Listeria vaccines. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
 CC on 29-AUG-2003 to standardise OS field)
 XX SQ Sequence 12278 BP; 4018 A; 2610 C; 1808 G; 3841 T; 0 U; 1 Other;
 Query Match 12.3%; Score 75; DB 6; Length 12278;
 Best Local Similarity 46.5%; Pred. No. 2.3e-13;
 Matches 280; Conservative 0; Mismatches 316; Indels 6; Gaps 1;
 QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTGGCGTTCGGATCTTATGGCAAGTCTC 60
 Db 9490 ATGATTGATTAGTTAATGTCAGTAAAGATACAAGATAAATGATTGTTGAAAANGTC 9431
 QY 61 AGTCATAAGTTTTTACCAGGAACATGACGACTGACTGAGCGTCCGGTTCAGGAAA 120
 Db 9430 TCTCTTCTTANTGGAGCTGGTGAATTTATTGAGTCGTCGGGAGAGTGGTAGTGGGAAG 9371
 QY 121 TCGACTTTCTCACTGCTTGGCCACCTTGCAAAACCAAGTTCCGGACAGATCCCTTCTC 180
 Db 9370 ACACGCTCTAATGTTATCGACACCTAGATTCAAAGATAGTGGGAAGTATTATT 9311

us-09-868-338-7_copy_1117_1725.rng

Mon Aug 2 09:36:54 2004

Job time : 455 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 10:28:19 ; Search time 99 Seconds
(without alignments) 3413.789 Million cell

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609
Sequence: 1 atgatagaatcaatgacct.....cggataccattatcaaaccta 609

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

- Job processing. Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Issued Patents NA:*

Issued Patents NA: *

1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*

/cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*

```
4: /cgn2_6/ptdata/2/ina/6B_COMB.seq:*
```

/cgn2_6/ptodata/2/ina/PCTUS_COMB.sec

```

e: /cgn2_6/ptodata/2/lna/backfiles1.se

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	75.4	12.4	21706	4	US-08-961-527-36	Sequence 36, Appl
2	74.6	12.2	580073	4	US-08-545-528D-1	Sequence 1, Appli
C 3	74.6	12.2	580073	4	US-08-545-528D-1	Sequence 1, Appli
4	71.8	11.8	813	4	US-09-107-532A-3565	Sequence 3565, Ap
5	71.6	11.8	1080	4	US-09-543-681A-1259	Sequence 1259, Ap
6	71.2	11.7	771	4	US-09-134-001C-1763	Sequence 1763, Ap
7	69.6	11.4	681	4	US-09-134-000C-2474	Sequence 2474, Ap
8	68.4	11.2	11050	4	US-08-956-171E-96	Sequence 96, Appl
9	66.6	10.9	1191	4	US-09-489-039A-3320	Sequence 3320, Ap
10	66	10.8	720	4	US-09-143-000C-3762	Sequence 3762, Ap
11	65	10.7	14872	4	US-08-961-527-72	Sequence 72, Appl
12	64.8	10.6	711	4	US-09-134-001C-1758	Sequence 1758, Ap
13	64	10.5	1149	4	US-09-134-001C-602	Sequence 602, App
C 14	63.8	10.5	4535	4	US-09-134-001C-602	Sequence 1758, Ap
C 15	62.6	10.3	4403765	3	US-09-103-840A-2	Sequence 187, App
C 16	62.6	10.3	4411529	3	US-09-103-840A-2	Sequence 2, Appl
17	61.6	10.1	765	4	US-09-134-001C-1603	Sequence 1, Appl
18	61.6	10.1	813	4	US-09-134-001C-2724	Sequence 1603, Ap
C 19	61	10.0	6593	4	US-08-961-527-195	Sequence 2724, Ap
C 20	61	10.0	8114	4	US-09-461-527-1528	Sequence 195, App
21	60.8	10.0	4669	2	US-09-583-702B-29	Sequence 29, Appl
C 22	60.8	10.0	20986	4	US-08-563-276-18	Sequence 29, Appl
C 23	60.2	9.9	1830121	4	US-08-961-527-54	Sequence 54, Appl
C 24	60.2	9.9	1830121	4	US-09-557-894-1	Sequence 44, Appl
C 25	60	9.9	1830121	4	US-09-643-990A-1	Sequence 1, Appl
C 26	59.6	9.8	10690	4	US-09-489-039A-5565	Sequence 1, Appl
C 27	59.2	9.7	2726	1	US-08-961-527-93	Sequence 5565, Ap
					US-08-961-527-93	Sequence 93, Appl
					US-08-961-823-1	Sequence 1, Appl

Query Match 12.4% Score 75.4 DB 4: Length 21706

Query Match	12.4%;	Score 73.4;	DB 4;	Length 21706;
Best Local Similarity	50.4%;	Pred. No. 1.7e-15;		

Matches	212;	Conservative	0;	Mismatches	206;	Indels	3;	Gaps	1;
---------	------	--------------	----	------------	------	--------	----	------	----

[illegible]

99 TGGAGCGTCCGGTTCAGGAAATCGACTTTGCTCAACTGTCTTGGCACACTTGACAAACC 158

TGGTGCCTTCAGGTGCGCAGCTCAACACTTTCCGCCAATCCC
14961 TGGTGCCTTCAGGTGCGCAGCTCAACACTTTCCGCCAATCCC

14961 TGGTGCTTCAGGTGCAGGCAAGTCAACAGTCTCTTAACCTTCTTGGGGGAATGGATACCAA 1502

QY 159 AAGTCCGGACAGATCCCTTTCGAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCA 218
DB 15021 TGATGAAGGGGAAATCTGGATTGATGGTGTAAATATTGCGGATATATAGTTCGCCACGCG 15080
QY 219 ACGGTTATACAGGAAATACGGTGGGCTATTTTATTTCAAGATTAATGCGCTTGATTCGCGA 278
DB 15081 CACCAATTTACCGTAGAATATGATGGGTTTGTTCATGTTTATTAATCTAGTTCTAA 15140
QY 279 CAGACAGTTAAATTCACCTTCAGCTTCGGTGGGAAACACAAATGCCCTGAAATTC 338
DB 15141 TCTCAGAGCTAAGAAATGTGAACCTGGCTTCTGAAATTTGACAGATGCGCTTGAATCC 15200
QY 339 ---TCAAGTACTTCATGCTTGGTCTTGAGTGGTTCGAGGAAAGCCAGTTTTTGAAC 395
DB 15201 TGATCAGGCTTGACAGATGATGGTCTGGCTCATCTCAATACTTCCAGGCCAGCT 15260
QY 396 CTCGGTGGGACACAAACGAACTGGTGGCGGGTACTGCTCAAAATCCCGAAT 455
DB 15261 TTCTGGAGGGAGCAACAGCGATCTCCATTCACCGCGGTAGGCCAAATCTTAAAT 15320
QY 456 AATTCGGCTGATGAACCAACCGAGCTCTAGATTTTACAAACAGTGAAGTCTATAGA 515
DB 15321 TCTCCTTTGTGATGAACCGAGCTGGAGCTTGGATTTATCAGACGGGCAAGCAGTTTGA 15380
QY 516 A 516
DB 15381 A 15381

RESULT 2
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match 12.2%; Score 74.6; DB 4; Length 580073;
Best Local Similarity 50.1%; Pred. No. 2.6e-14;
Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;

QY 100 GGAGCGTCCGGTTCAGGAAATCGACTTTGCTCACTGCTTGGCAGACACTTGGCAACCA 159
DB 78490 GGCGCTTCAGGATCTGGTAAACCAATCTACTAAACATTAATTCAGGATGATAGACT 78549
QY 160 AGTTCGGACAGATCCCTTGTTCAGGATGTAGACCTTCTGAACTCTCTACGCGTAAGCAA 219
DB 78550 TCTAGTGGTAGCTTATGTCTCAATGGTTATACATGATTTGTTAAATGATAGAAGCTC 78609
QY 220 CGGTTATACAGGAAATACGGTGGGCTATTTATTTCAAGATATGCTTGAATCCCGAC 279
DB 78610 ACTAAATTCGCTCAAAAGTATGTGTGTATCTCTTCAACATATCGGTTTATACCTAAT 78669
QY 280 AGGACAGTTAAATTCACCTTCAGCTTGGGTTGGGTAAGGACAC- - - - -AAATGG 327
DB 78670 TTAACAGTTAGAGAAACATTTGATAGAGGCAATCTTCAACAGATCCTTAGTAAAGG 78729

QY 328 CCTGAATTCCTCAAGTACTTCTATGCTGTGGTCTTGGTCTTGGAGTAAAGCCAGTT 387
DB 78730 ATCAGCATTTGATGCCATTTTAGAAGCGGTGGGATAGTTTGCATAAAGAGCTTCCT 78789
QY 388 TTTGAATCTCTCTGGTGGGCAACAAACGAACTGCTTGGCCCGGCTACTCTCAAAAT 447
DB 78790 AATGAATTCAGTGGTGGGCAACAGCAACGCTGTTCCATTCGAAGAGCTTTTGGCTAAAC 78849
QY 448 CCGCGAATATTTCTGGCTGATGACCAACCGGAGCTCTAGATTTTAAACAAACAGTGACTA 507
DB 78850 CCGTTAATTTTGGTGGTGAACCTACTGCGGCACTTGATCTTGAGATGACCCAAAT 78909
QY 508 GTCATAGAAGCACTTGAGAGCACT 530
DB 78910 GTTTTAAACAGTTTTTAGCAAT 78932

RESULT 3
US-08-545-528D-1/c
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match 12.2%; Score 74.6; DB 4; Length 580073;
Best Local Similarity 50.1%; Pred. No. 2.6e-14;
Matches 216; Conservative 0; Mismatches 209; Indels 6; Gaps 1;

QY 67 AAGTTTTCACAGGAACTATGACAGCACTGACTGAGCGTCCGGTTCAGGAAATCGACT 126
DB 576972 AAAATCTCCAGGAGAAATTTGTTTATTTCTGTAATCTGTTAGTGTAAACCCAGC 576913
QY 127 TTGCTCAACTGCTTGGCAGACTTGCACAAACCAAGTTCGGACAGATCTTGTGAGGAT 186
DB 576912 TTACTTCTTTAAATTCGCGATTAGATCGACCAACTAGCGCGGATGTTTGTGTTGGA 576853
QY 187 GTAGACCTTCTGAAACTCTCTACGCGTAAGCAACGTTATACAGGAAATATACGTCGGC 246
DB 576852 ACTAACCAATATGTTGTAGTGATGCAAACTAACCGGTTAAGAAATAAAACGTTGGT 576793
QY 247 TATTTATTTCAAGATATGCTTGAATTCGCCAGACAGGACAGTTAAATTCACCTTCAGCT 306
DB 576792 TATATTTTCAACATATGTTTATTTGCGTATCTTGACGTTGATGATAATATAAGCTT 576733
QY 307 GCGGTGGGAAACACAAATGGCCTGAAATTCCTC- - - - -AAGTACTTCATGCTGTTGGT 360
DB 576732 GCTCTACCATTAATAAAGGTTTAAATAATAATCTTGAAGAAATTTATAGAACGCTTGAA 576673
QY 361 CTTGAGTGGTTCGAGGAAAGCCAGTTTTCAGCTCTCTGTCGCGACCAACCAAGCACT 420
DB 576672 TTAAGAAGCAATAGACATAGAAAGTTCAATACTCTCTGTTGACCAACCAACGCTGT 576613
QY 421 GCGTTGGCGGGTACTCTCTCAAAATCCCGAATAATTTGCTGATGATGAACCAACCGGA 480
DB 576612 GCTATTGCTCGGCTTTAATTAAGAAGCACTTAAATTTGTTGTTGATGAACCTACTGGA 576553

QY 481 GCTCTAGATT 491
Db 576552 GCAGTAATAT 576542

RESULT 4
US-09-107-532A-3565
; Sequence 3565, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinietello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3565:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 813 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3) LOCATION 1...813
; SEQUENCE DESCRIPTION: SEQ ID NO: 3565:
US-09-107-532A-3565

Query Match 11.8%; Score 71.8; DB 4; Length 813;
Best Local Similarity 49.6%; Pred. No. 4.1e-15;
Matches 212; Conservative 0; Mismatches 212; Indels 3; Gaps 1;

QY 99 TGGAGCGTCGGTTTCAGGAAATCGACTTCTCAACTGCTTGGGACACTTGCAGAAACC 158
Db 228 TGGTCCAGTGGACGAGGAAGTCAACGATCTCTTAATTTAGTGGGATGGATACACC 287

QY 159 AAGTTCGGGACAGATCCTTGTTCGAGGATGTAGACTTCTGAAACTCTCTACGGGTAGCA 218
Db 288 TGATCAAGGTGAGATCAATATTGATGACAGATATTGGCAATTTTCAGATTAACAGCT 347

QY 219 ACGGTTATACAGGAAATACGGTGGGCTATTTATTTCAAGATATGCTTGTATCCCGA 278
Db 348 TACTGCTTACCGACGACAGATGTAGGTTTCGTTTCCCAATTTCTACAATCTCGTTCCTAA 407

QY 279 CAGGACAGTTAAATTCACCTTCAGCTTGGGTGGAAAAA---CACAAATGGCTGAAT 335
Db 408 TCTAACAGCAAAAGAAATATGTGAATTCGCAACGGAAGTTTCTCTGTCGCTTGTATCC 467

QY 336 TCCTCAAGTACTTCATGCTGCTTGTGCTTGTAGTCTGTCGAGGAAAGCCAGTTTGAAT 395
Db 468 TGTAGAGGTATTGAGGCAAGTGGGCCCTAGCTCATCGATTGAATAACTTTCCGTCCTCAAT 527

QY 396 CTCTGGTGGGAAACAACAAGAACTCGTTGGCCCGGTACTGCTCAAAAATCCCGAAT 455
Db 528 ATCAGAGGGGAACAGCAGAGGGTTTCAATTTGCTGCTGACTAGCTTAAAAATCCCAACT 587

QY 456 AATTCGTGCTGATGAACCAACCGAGCTCTAGATTTTAAACAACAGTGTAGTCTATAGA 515
Db 588 TCTTTTATGTGATGAACCGACAGCGCGTGGATTTTGAACGGGCAACGAGTATTGAA 647

QY 516 AGCATTG 522
Db 648 ATTATTG 654

RESULT 5
US-09-543-681A-1259
; Sequence 1259, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1259
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1259

Query Match 11.8%; Score 71.6; DB 4; Length 1080;
Best Local Similarity 47.4%; Pred. No. 5.8e-15;
Matches 215; Conservative 0; Mismatches 239; Indels 0; Gaps 0;

QY 46 TTATGCAAGGTCTCAGTCATAAGTTTTTACAGGAACAATGACAGACTGACTGGAGCG 105
Db 52 TTGTGCTGGAAGTGAATACTCAATTACCAACAGAAAGTATCACGCGGTATTTGGCTTA 111

QY 106 TCGGTTTCAGGAAATTCGACTTTGCTCAACTGCTTTGGCACACTTGACAAACCAAGTTCC 165
Db 112 TCTGGCGCAGGTAAACGTCGCTCAATATGATGATTTGGTTTAAACCAACCTGATTCA 171

QY 166 GGACAGATCCTTGTTCGAGGATGTAGACTTCTGAAACTCTCTACGCGTAAGCAACGGTTA 225
Db 172 GGACATATTGTGTTAAATAATCACACATTTGGTGGATATTGAGAAGAAAATTTATCTGCCA 231

QY 226 TACAGGAAATATACGCTGGCTTATTTTTCAGGATTTATGCTTATTCGCGACAGGACA 285
Db 232 CCCGAAAGCGCAAGTAGGCTATGCTATTTCAAGTGCACACTATTTCTCTCATTTATACC 291

QY 286 GTTAAATTCACCTTCAGCTTTCGGTGGGAAACCAAAATGCGCTGAAATTCCTCAAGTA 345
Db 292 GTAAAGGAAATTTACTTTAGCTATGTCCACCAATGAAAGTACAGTTTGTATGAAT 351

QY 346 CTTTCATGCTGTGGTCTTGTAGTCTGTCGAGGAAAGCCAGTTTGTGAATCTCTGCTGGC 405
Db 352 ATTTATCTATTAGGTATCGAACATCTACTTTTCGGGTTTTTCCTATTTACGCTATCTGCTGT 411

QY 406 GAACAACAACGAACTGCGTTGGCCCGGTACTGCTCAAAAATCCCGCAATAATTCGCTGCT 465
Db 412 GAAACACACGCTGTGGCTATTGGAGTGCATTTAATTAAGTGCACCGCATTTATGCTAATG 471

FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2474
LENGTH: 681
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-134-000C-2474

Query Match 11.4%; Score 69.6; DB 4; Length 681;
Best Local Similarity 49.4%; Pred. No. 2.2e-14;
Matches 251; Conservative 0; Mismatches 239; Indels 18; Gaps 2;

QY 77 CAGGAACAATGACAGCAGTCTGAGGCTCGGGTTCAGGAAATCGACTTTGCTCAACT 136
Db 95 CTGGAGAATTTTATGCGATTATGGGTCTCTGGTCTCGGAAATCGACTTTGATTAAT 154
QY 137 GTCTTGGCACACTTGCACAAACCAAGTTCGGGACAGATCCTTGTCCAGGATGTAGACCTTC 196
Db 155 TGTAGGCTTTATGATAAAAGTTTGAAGGAACCTATTATTTCAGAGACCGTGAATTTG 214
QY 197 TGAACCTCTACGGGTAAAGCAACGGTTATACAGGAAATACGGTGGCTATTTATTTTC 256
Db 215 GCGACTTTTCTGATAAGGAACCTATCCCGAATTTAGAAATGAAGCAGTCCGGCTTTGCTTTTC 274
QY 257 AAGATTATCGCTTGTATTCGACAGGACAGTAAATTTCAACCTTCAGCTTTCGGGT----- 311
Db 275 AGAATTTTATTTAGTTTAAATGAAACACTAACAGTTGAGAAATCAATTGATTTACCTTTTAT 334
QY 312 -----GGAAAAACAATGCGCTGAAA-----TTCTCAAGTACTTTCATGCTGTG 358
Db 335 ATAGTGGGTTAAACCCCAAGGAAGCAAGATCGTGTCCACGAAGTTCTGACAAAAGTCG 394
QY 359 GCTTTCAGTCTGTCGAGGAAAGCCAGTTTGTGAACCTCTCTGTCGCGAACAACAAACGAA 418
Db 395 GCTTTCAGATAAAGCAAGAGAGATCCAAACACATTTTCGGCGGACAAACAACGATG 454
QY 419 CTGCGTTGGCCCGGCTGCTCAAAAATCCCGCAATTAATCTGGCTATGAAACCAACCG 478
Db 455 TTCCCATTCGCGTGGGATTTGCAATCGGCCAAGTTTATTATTGAGATGAACCAACTG 514
QY 479 GAGCTTAGATTTTAAACACAGTGTAGCTAGTATCATAGACATTGAGAGCACTCGCGACA 538
Db 515 GGCCTTTAGATAGTAAACATCGGAAGAAATTTTAAGCTCTTTTCAGCAATTAACAATG 574
QY 539 AAGCGCGCCACCGTTGTTGTTGCTACGCA 566
Db 575 AAGGTGTCACCATTTATTTAGTACACA 602

RESULT 8
US-09-134-000C-2474
Sequence 96, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA


```

; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2762
Query Match 10.8%; Score 66; DB 4; Length 720;
Best Local Similarity 55.8%; Pred. No. 4.4e-13; Indels 0; Gaps 0;
Matches 126; Conservative 0; Mismatches 100;

QY 341 AAGTACTTTCATGCTGCTTGGTCTTCTGAGTCTGTTTCGAGGAAAGCCAGTCTTGAACCTCTCTG 400
Db 413 AATACCTAAAATTTGTCGGTTTGGAAACAAAAGCCAAATCAAAAGTACTGAACTTCTG 472
QY 401 GTGGCGAACACAGCACTGCTTGGCCGGTACTGCTCAAAAATCCCGAATAATTC 460
Db 473 GAGGACACCAACAAAGATGACCAATGCTGCTTCTTGGTGAATGATCCATATTGATTA 532
QY 461 TGGCTGATGAACCAACCGGAGTCTTAGATTTTAAACAACAGTGAAGTACTAGTATAGAACAT 520
Db 533 TGGCTGATGAACCGACTGGGCGATTTGGATAGTGAACGGGTACAGAAATATGGAATTAT 592
QY 521 TGAGAGCACTCGCCACAAAAGCGCCACCGTGTGTTGTTGCTACGCA 566
Db 593 TGAAGAATTGAATGAACAGGGCAAAACAATTTGTCATGTTGACACA 638

RESULT 11
US-08-961-527-72
; Sequence 72, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-72
Query Match 10.7%; Score 65; DB 4; Length 14872;
Best Local Similarity 49.1%; Pred. No. 6.7e-12; Indels 18; Gaps 2;
Matches 244; Conservative 0; Mismatches 235;

QY 91 GCAGTACTGGAGCGTCCGGTTCAGGAAATCGACTTTGCTTCAACTGCTTGGCACACTT 150

```

```

Db 207 GCCATCATGGTGAGTCTCGTTTCTGGTAAATCAACTCTTCTCAATATTCTAGCTAATGTTG 266
QY 151 GACAAACCAAGTTCCGGACAGATCTTGTGCGAGGATGTAGACCTTCTGAAAACCTCTCTACG 210
Db 267 GATAAACCAAGTCGTGGTTCAGGTTTACITTAATGAATGAACTGACACCGCACTATTAAAT 326
QY 211 CGTAAGCAACGGTTATACAGGAAATAACGTTGGGCTATTATTATTCAGATATTACCTTGG 270
Db 327 TCACAGGCTTCTAGTTTCCGGCGTGAAAGAGCTAGGATTGTCTTCCAAGACTTTAACTTG 386
QY 271 ATTCCGACAGACAGTGTAAATTTCAACCTTCAAGTTGGGTGCGTGAATAACACAAATGGCT 330
Db 387 CTAGATACTCTGTTGTTAAGACAAATATCTTGGCTTCCGTTGTCTTGCAGAGACCT 446
QY 331 GAAAT-----TCTCAAGTACTTTCATGCTG-----TTGGTCTTGAAGTCGTTTC 372
Db 447 ATAACGGAGATGATGAAGAAATTTGGTGTGACAGCTGAGAAATCTGGGTATTAAACCAATTG 506
QY 373 GAGGAAAACCGAGTTTGTGAACCTCTCTGTGCGCAACACACAGAACTGCGTTGGCCCGG 432
Db 507 CAAGAGAGTACCCCTTACGAGATTCTGTGTGTGCAAAACAGCGTGTAGCAGTAGCCCGC 566
QY 433 GTACTGCTCAAAAATCCCGAATAATTCTGGTGTATGAACCAACCGAGCTCTAGATTTA 492
Db 567 GCCATCATCACAGAACTGAAATTTCTCTTGGGACGAGCCAAACAGGAGCCCTTGAATCC 626
QY 493 ACAACAGTGAAGTCTAGTATAGAACATGAGAGCACTGCGCGACAAAGGCGCCACCGTT 552
Db 627 AAGTCATCTGCAGCCTTACTTGTATGCTTTAATGAATCAATGAGCGTGGGCAACCAATC 686
QY 553 GTGTGTTGCTACGCACTC 569
Db 687 CTCAGTGAACCCACTC 703

RESULT 12
US-09-134-001C-1758
; Sequence 1758, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1758
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1758
Query Match 10.6%; Score 64.8; DB 4; Length 711;
Best Local Similarity 47.2%; Pred. No. 1.2e-12;
Matches 250; Conservative 0; Mismatches 262; Indels 18; Gaps 1;

QY 47 TATGCGAAGTCTCAGTCAATAGTTTACCAGGAAACATGACAGCACTGACTGAGCGT 106
Db 83 ATTTGAAAGACATAGATATTCATATGATGAGGTGAATTTATTGCTATCATGGTCCGT 142
QY 107 CCGGTTTCAGGAAATCGACTTTGTCTCAACTGTCTTGGCACACTTTCACAAACCAAGTTCCG 166
Db 143 CTGGTTCAGTAAAGTACGTTAATTAATCTTGGGATTTATTGATCGTGCATATGAAG 202
QY 167 GACAGATCCTTGTGCGAGATGTAGACCTTCTGAACACTCTTACCGGTAAAGCAAGTTAT 226
Db 203 GAGAATACTTTTTTAAACAATGAGATTTATCAAAAAGCTCAGATAATAAGCTCCGAGAAA 262

```

QY 227 ACAGGAAAAATACGGTGGGCTATTATTTCAAGATTATCCCTTGATTCCCGACAGACAG 286
Db 263 TTCCGAATCATACTGTAGTGTTCGTTCTTTTCAGAAATTTTAAATTTGATTCAAAATAACACTA 322
QY 287 TTAATTTCAACTTTCAGTTCGGTGG-----AAAAACAACAATGCG 328
Db 323 TTTTGAATAATGAAGTATTCGCTAATTTATTAATGTTTGTAGTAAATAAGCGAGAAAA 382
QY 329 CTGAAATTCCTCAAGTACTTCATGCTGTGTGCTTCGAGTCGTTCGAGGAAAAAGCGATTT 388
Db 383 GCAAGTCTTTAGACGACTTCATGCTCGGTCTAAAGGTAAAGAGAAATCTCTTACCAA 442
QY 389 TTGAATCTCTGTCGCGGACACACACGAGTGGTTCGCGGCTGCTCAAAATC 448
Db 443 ATAAATTTGTCAGAGGCGGACACACGCTGTAGCTATTGCGCGGCAATTTAATATGACC 502
QY 449 CCCGAATAATTTCTGGTGTATGAACCAACCGGAGCTCTAGATTTAAACACAGTGTAG 508
Db 503 CTAAGTTTATTTGCGGTGACCTACTTGGGCTCTTGATTCGAAGACATCTCAAGATA 562
QY 509 TCATAGAGCTTGTAGAGACTCGCGACAAAGGCCCGCCCGTGTGTGTT 558
Db 563 TTATGAGCTTTTCGTAAACTGAAACAAAGACAGAATACAACTATATT 612

RESULT 13

US-09-134-001C-602

; Sequence 602, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: LYNN DOUCETTE-STAMM ET AL

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 602

; LENGTH: 1149

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-602

Query Match 10.5%; Score 64; DB 4; Length 1149;
Best Local Similarity 45.2%; Pred. No. 3e-12;
Matches 235; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

QY 11 TCAATGACCTCAAGAAATCTTTGGCGTTCGGATCTTATGGCAAGTCTCAGTCATAAGT 70
Db 74 TTAAGATGTCAGTAAGGCTTTGAAGATGTACAATACTAAATGAAATTAATTGATA 133
QY 71 TTTTACAGAAACAATGACAGACTGACTGAGCGTCCGGTTCAGGAAATCGACTTGC 130
Db 134 TTGAACCCAGGCTATTTTATACACTATTAGTCCCTCAGGTTGTGGAACCAACAATTT 193
QY 131 TCAACTGCTTGGCACACTTGACAAACCAAGTTCGAGACAGATCTTGTGAGGATGTAG 190
Db 194 TAAACTCATAGCAGGATTTGAATATCCGATAGTGAGATATATATAAGATAAAC 253
QY 191 ACCTTCTGAACTCTCTACGCGTGAAGCAACGGTTATACAGGAAAAATACGGTGGGTATT 250
Db 254 CTAATGTTAAATGCCACCAGTAAGCGTAAGTAAATCTGATTTCCAAGACTATGCAT 313
QY 251 TATTTCAAGATTATGCTTGNATCCCGACAGGACAGTAAATTCACCTTCAGCTTCCGG 310
Db 314 TGTTTCCACATTTAAATGTATGACAATATGCTATGTTTAAATTAATAAATAA 373
QY 311 TGGAAAAACAAATGGCCTGAAATTCCTCAAGTACTTCATGCTGTGTGCTTGTAGTCGT 370

Db 374 GTAAGTCAGAAATTAACGGTAAGGTTACTGAAGCACTTCAGTTGGTGAATTAAGTGGTT 433
QY 371 TCGAGGAAAAAGCCAGTGTGTTTGAACCTCTCTGTCGCGGAAACAAACAGAACTGCGTTGCCCC 430
Db 434 ATGAACATAGGCAATACAAGGTATGAGTGTGGACAAAAACACACGTTGTAGCCATAGCAC 493
QY 431 GGGTACTGCTCAAAATCCCGAATATTTCTGGCTGATGAACCAACCGGAGCTCTAGATT 490
Db 494 GGGCAATGTTAATGAGCCTGAAATATATTATTAGATGAGTCTTTATCGCATTAGATT 553
QY 491 TAAACAACAGTGTAGCTAGTCTCATAGACATGAGAGCACT 530
Db 554 TAAATTAACGAACCTGAAATGCAATATTTATTGAGAGAACT 593

RESULT 14

US-09-221-017B-187/c

; Sequence 187, Application US/09221017B

; Patent No. 6444799

; GENERAL INFORMATION:

; APPLICANT: ROSS, BRUCE C.

; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA: US/09/221,017B

; FILING DATE: 23-DEC-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1182

; FILING DATE: 31-DEC-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1546

; FILING DATE: 30-JAN-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP2911

; FILING DATE: 09-APR-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/AU98/01023

; FILING DATE: 10-DEC-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: MOIRCV, Gladys H

; REGISTRATION NUMBER: 32,430

; REFERENCE/DOCKET NUMBER: 27340-20021.00

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-813-5600

; TELEFAX: 650-494-0792

; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 187:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4635 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: UNKNOWN

; ORIGINAL SOURCE:

; ORGANISM: PORYPHYROMONAS GINGIVALIS

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 1...4635

US-09-221-017B-187

Query Match 10.5%; Score 63.8; DB 4; Length 4635;
Best Local Similarity 55.7%; Pred. No. 8.5e-12;
Matches 122; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 390 TGAATCTCTGTCGCGGACACCAACCGGAGCTCTAGATTAAACAAACAGTGTAGTGT 509
DB 1674 TCAGCTCTCCGAGACACCAACCGGAGCTCTAGATTAAACAAACAGTGTAGTGT 1555
QY 450 CGAATAATTCGTCGATGAACCAACCGGAGCTCTAGATTAAACAAACAGTGTAGTGT 509
DB 1614 GAAGCTCATCTCGCGGATGAACCAACCGGAGCTCTAGATTAAACAAACAGTGT 1555
QY 510 CATAGAGCATGTAGAGCACTCGCGGACCAACCGGAGCTCTAGATTAAACAAACAGTGT 569
DB 1554 CATGGAATCTGTCAGAGCTCTCAATCGGAGGTGCAACCATCGTCTATGTTGTCAGCACTC 1495
QY 570 GCCCTCTTCCGAGATCAGCGGATACCATTTATCAAACT 608
DB 1494 CGAGCAGATGCAGTAGTCCGCGCGCATCATCT 1456

RESULT 15

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 10.3%; Score 62.6; DB 3; Length 4403765;
Best Local Similarity 53.0%; Pred. No. 1.7e-09;
Matches 134; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 335 TTCCTCAAGTACTTTCATGCTGTGCTCTGAGTGTGTCGAGGAAAGCCAGTGTTCGAAC 394
DB 3466760 TCCCGGAGTGTCTGAGACGGTGTGTCGCGCAGGCGCAATCGGCTCCAGACGAGC 3466701
QY 395 TCTCTGTGGCGAACAACAAGCACTCGGTTGGCCCGGTAAGTGTCAAAAATCCCGGAA 454
DB 3466700 TGTCCGCGCGGAGCAACAGCGGTTGCGATCGCCCGCGCTTTGTCAACCGCGCGCTGG 3466641
QY 455 TAATTCGCTGATGAACCAACCGGAGCTCTAGATTAAACAAACAGTGTAGTGTATAG 514
DB 3466640 TACTGTGCGCGAGCAGCCCAACCGGAACTCGACCGGAGACGATGAGGACATATGG 3466581
QY 515 AAGCAATTGAGAGCACTCCCGGACAAAGCGCCACCGTGTGTTGTTGTCAGCACTCCGCC 574
DB 3466580 ACTTATTGGAGCGGATCAACCGGACCGGACGAGGTGCTGATGGCCACGACGACACC 3466521
QY 575 TCTTCGCGAATC 587
DB 3466520 ACATCGTCGACTC 3466508

Search completed: July 30, 2004, 14:16:21
Job time : 111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 13:27:50 ; Search time 506 Seconds
(without alignments)
5901.210 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609

Sequence: 1 atgatagaatcaatgacct.....cggataccattatcaacta 609

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2.*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80.8	13.3	684707	16	US-10-398-221-9
c	80.8	13.3	3011208	16	US-10-398-221-2058
2	77.4	12.7	711	9	US-09-815-242-9384
3	75.4	12.4	21706	13	US-10-158-844-36
c	75	12.3	12278	16	US-10-398-221-3886
4	74.6	12.2	1401	13	US-10-282-122A-27244
5	74.6	12.2	31241	10	US-09-960-870-13
6	74.6	12.2	31241	10	US-09-960-858-13
7	74.6	12.2	31241	10	US-10-251-668-13
8	74.6	12.2	580073	15	US-10-205-220-1
9	74.6	12.2	580073	15	US-10-205-220-1
c	74.6	12.2	580073	15	US-10-282-122A-34668
10	74.6	12.2	580073	15	US-10-282-122A-34668
11	74.6	12.2	580073	15	US-10-282-122A-34668
12	73.6	12.1	5312	16	US-10-398-221-3820
13	73.6	12.1	5312	16	US-10-398-221-3820
14	73.2	12.0	684	13	US-10-282-122A-30817

15	72.4	11.9	666	9	US-09-738-626-956	Sequence 956, App
16	72.4	11.9	789	13	US-10-627-476-441	Sequence 441, App
c	72.4	11.9	3309400	9	US-09-738-626-1	Sequence 1, Appli
17	71.6	11.8	1062	13	US-10-282-122A-32942	Sequence 32942, A
18	71.2	11.7	319630	16	US-10-398-221-7	Sequence 7, Appli
19	69.6	11.4	32768	9	US-09-070-927A-123	Sequence 123, App
20	68.6	11.3	630	15	US-10-307-234-12	Sequence 12, Appli
21	68.6	11.3	7000	15	US-10-307-234-3	Sequence 3, Appli
22	68.4	11.2	11050	8	US-08-781-986A-96	Sequence 96, Appli
23	68.4	11.2	11050	13	US-10-329-624-96	Sequence 6038, Ap
24	68.2	11.2	687	9	US-09-815-242-6038	Sequence 238, App
25	66	10.8	29729	9	US-09-070-927A-238	Sequence 10135, A
26	65.6	10.8	1149	13	US-10-282-122A-10135	Sequence 27896, A
27	65.2	10.7	1398	13	US-10-282-122A-27896	Sequence 24170, A
28	65	10.7	726	13	US-10-282-122A-24170	Sequence 3721, Ap
29	65	10.7	2293	16	US-10-398-221-3721	Sequence 72, Appli
30	65	10.7	14872	13	US-10-158-844-72	Sequence 68, Appli
31	64.6	10.6	8921	9	US-09-070-927A-68	Sequence 187, App
32	63.8	10.5	4635	13	US-10-194-163-187	Sequence 1791, Ap
c	63.6	10.4	732	16	US-10-398-221-1791	Sequence 6, Appli
33	63	10.3	213251	16	US-10-398-221-6	Sequence 2058, Ap
34	62.6	10.3	3011208	16	US-10-398-221-2058	Sequence 9252, Ap
35	62.2	10.2	1163020	16	US-10-398-221-10	Sequence 3588, Ap
36	62.2	10.2	816	9	US-09-815-242-9592	Sequence 23957, A
37	62.2	10.2	816	9	US-09-815-242-9592	Sequence 2421, Ap
38	62.2	10.2	816	9	US-09-815-242-9592	Sequence 195, App
39	62.2	10.2	3438	16	US-10-398-221-3588	Sequence 29, Appli
c	62	10.2	1059	13	US-10-282-122A-23957	Sequence 342, App
40	61.6	10.1	1059	13	US-10-282-122A-23957	
41	61.6	10.1	577	9	US-09-974-300-2421	
42	61.4	10.1	577	9	US-09-974-300-2421	
43	61	10.0	6693	13	US-10-158-844-195	
c	61	10.0	8114	15	US-10-114-170-29	
44	61	10.0	8114	15	US-10-114-170-29	
c	60.8	10.0	741	10	US-09-769-787-342	

ALIGNMENTS

RESULT 1

US-10-398-221-9
; Sequence 9, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 684707
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-9

Query Match 13.3%; Score 80.8; DB 16; Length 684707;
Best Local Similarity 47.2%; Pred. No. 5.3e-14;
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;

QY 1 ATGATGAAATCAATGACCTCAAGAAATCTTTGGGCTTCGGATCTTATGCGAAGTCTC 60
Db 346005 ATGATGAAATAGTAATGTGCTAGTAAAGATCAAGATAAATGATTTTGGAAAAAGTC 346064
QY 61 AGTCATAAGTTTATACAGGAACAATGACAGCACTGACTGCGGCTTCAGGAAAA 120

Db 346065 TCCTCTTCTATTGGAGCTGGCGAATTTATTGTCAGTCTGCGGAGAGTGGTAGTGGGAAG 346124
QY 121 TCGACTTTGCTCAACTGCTTGGCAGACTTGCACAACTTGCAGACAGATCCTTTGTC 180
Db 346125 ACAACGCTGCTAAATGTTATCGGACACTAGATTCAAAAGATAGTGGCAGATTATTT 2934630
QY 181 GAGGATGTAGACCTTCTGAACTCTCTACGCGTAAAGCAACGGTTATACAGGAAAAATACG 240
Db 346185 AACGAGATGGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAAAAAAGAGGTG 2934576
QY 241 GTGGGCTATTATTTCAGATTATGCTTCCGACAGGACAGTTAAATTCACCTT 300
Db 346239 TTAGGTTTATATTCCAAATTTCTATTGATGGAATGAAACAGTGTAGAAAACTTA 2934516
QY 301 CAGCTTGGCGTGGAAAAACACAAATGGCTGAAATTCCTCAAGTACTTTCATGCTGTTGGT 360
Db 346299 TCCATTACAGGTGGGAAAAATCGCAAGCTGATGATAGAGCATTTGGGAAGTAGGAATG 2934456
QY 361 CTGAGTCTCTCGAGAAAGCCAGTTTTCGAGTCTCTCGTGGCGAACCAACCACT 420
Db 346359 GATGAGAGCTATTAGCAAAAAAGTATACCAATTAAGTGGTGAGAAAAAACAACGGATT 2934396
QY 421 GCGTTGGCGCGGTACTGCTCAAAATCCCGAATTAATTCGGTGTGATGAACCAACCGGA 480
Db 346419 GCTATTGTGGCATTTTACTCAACCATTTCACTTTTACTTGGGAGCAACCACTGGC 2934336
QY 481 GCTCTAGATTAAACAAACAGTGGAGTGTAGTATAGAGCACTTGGCGACAA 540
Db 346479 AATTAGATGATAAAAAACAAATAATCATTAATTTAGTGGTGGAAAAAGCAA 2934276
QY 541 GGCGCCACCGTGTGTTGCTACGCACTGGCGCTTCTCCGAGAAATCAGCGGATACCAT 600
Db 346539 GGTAAGACTCGTTGTGTGTCAGCATGACCGGAAATATCAGGAAAAAGCATCGGGTC 2934216
QY 601 AT 602
Db 346599 AT 346600

RESULT 2

US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
US-10-398-221-2058

Query Match 13.3%; Score 80.8; DB 16; Length 3011208;
Best Local Similarity 47.2%; Pred. No. 1.2e-13;
Matches 284; Conservative 0; Mismatches 312; Indels 6; Gaps 1;
QY 1 ATGATAGAAATCAATGACCTCAGAAATCTTTGGCGTTCGGATCTTATGGCAGGTCTC 60
Db 2934809 ATGATTGAATAGTTAAATGTCTAGTAAAAAGATACAGATAATTTGTAATTTGGAAAAAGTC 2934750
QY 61 AGTCATAAGTTTTTACCAGGAAACATGACAGCACTGACGTGAGCGTCCGGTTCAGGAAA 120
Db 2934749 TCTCTTCTATTGGAGCTGGCGAATTTATTGTCAGTCTGGCGAGAGTGGTAGTGGGAAG 2934690

QY 121 TCGACTTTGCTCAACTGCTTGGCAGACTTGCACAACTTGCAGACAGATCCTTTGTC 180
Db 2934689 ACAACGCTGCTAAATGTTATCGGACACTAGATTCAAAAGATAGTGGCAGATTATTT 2934630
QY 181 GAGGATGTAGACCTTCTGAACTCTCTACGCGTAAAGCAACGGTTATACAGGAAAAATACG 240
Db 2934629 AACGAGATGGAATATCAGA-----CGAAAAAGAGGTTATGACTCTAAAAAAGAGGTG 2934576
QY 241 GTGGGCTATTATTTCAGATTATGCTTCCGACAGGACAGTTAAATTCACCTT 300
Db 2934575 TTAGGTTTATATTCCAAATTTCTATTGATGGAATGAAACAGTGTAGAAAACTTA 2934516
QY 301 CAGCTTGGCGTGGAAAAACACAAATGGCTGAAATTCCTCAAGTACTTTCATGCTGTTGGT 360
Db 2934515 TCCATTACAGGTGGGAAAAATCGCAAGCTGATGATAGAGCATTTGGGAAGTAGGAATG 2934456
QY 361 CTGAGTCTCTCGAGAAAGCCAGTTTTCGAGTCTCTCGTGGCGAACCAACCACT 420
Db 2934455 GATGAGAGCTATTAGCAAAAAAGTATACCAATTAAGTGGTGAGAAAAAACAACGGATT 2934396
QY 421 GCGTTGGCGCGGTACTGCTCAAAATCCCGAATTAATTCGGTGTGATGAACCAACCGGA 480
Db 2934395 GCTATTGTGGCATTTTACTCAACCATTTCACTTTTACTTGGGAGCAACCACTGGC 2934336
QY 481 GCTCTAGATTAAACAAACAGTGGAGTGTAGTATAGAGCACTTGGCGACAA 540
Db 2934335 AATTAGATGATAAAAAACAAATAATCATTAATTTAGTGGTGGAAAAAGCAA 2934276
QY 541 GGCGCCACCGTGTGTTGCTACGCACTGGCGCTTCTCCGAGAAATCAGCGGATACCAT 600
Db 2934275 GGTAAGACTCGTTGTGTGTCAGCATGACCGGAAATATCAGGAAAAAGCATCGGGTC 2934216
QY 601 AT 602
Db 2934215 AT 2934214

RESULT 3

US-09-815-242-9384
; Sequence 9384, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Hasselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9384

LENGTH: 711
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(711)
US-09-815-242-9384

Query Match 12.7%; Score 77.4; DB 9; Length 711;
Best Local Similarity 50.8%; Pred. No. 1.4e-14;
Matches 236; Conservative 0; Mismatches 226; Indels 3; Gaps 2;

Qy 55 GGCTCAGTCATAGTTTACACAGGAACAATGACAGCACTGACTGGAGCGTCGGGTCA 114
Db 82 GATCTGAATTTGAGATTGAAGGGGAACTGTTATTATCTTGTGCTTCAGGTGCA 141
Qy 115 GGAATAATGCACTTGTCTCAACTGCTTTGGCACACTTGACAAACCAAGTTCGGCACAGATC 174
Db 142 GCGAAGTCAACAGTCTCTTAACCTTCTTGGGGGAATGATACCAATGATGAAGGGGAAATC 201
Qy 175 CTTCGCGAGATGAGACCTTCTGAACTCTCTACGGGTAAAGCAACGGTTTATACAGGAA 234
Db 202 TGGATTGATGGTGTAATATTGCGGATATAGTTCCACAGCGCAACCAATTACCGTAGA 261
Qy 235 AATACGGTGGGCTATTATTTCAGATTATGCGCTTGATCCCGACAGACAGTTAAATTC 294
Db 262 AATGATGGGGTTGTTTTCAGTTTATAATCTAGTTTCTAATCTGACAGCTAAGGAA 321
Qy 295 AACCTTCAGCTTGGGTGGGAAA--AACACAAATGGGCTGAAAT--TCCTCAAGTACTTCAT 351
Db 322 AATGTGAACCTGGCTTCTGAAATTTGACAGATGCGCTTGAATTTCTGATCAGGCTTTGACA 381
Qy 352 GCTTTCGCTTGTAGTCTGTCAGGAAAAGCCAGTTTTCGAACTCTCTGCTGGCGGAACAA 411
Db 382 GATGTAGCTGGCTCATCTGCTAAATACTTTCCAGCCCGAGCTTCTGAGGGAGCAA 441
Qy 412 CAACGAATGCTGTTGGCGGGTACTGCTCAAAATCCCGAATAATCTGGCTGATGAA 471
Db 442 CAGCGAGTCTCCATTGCAACGCGGTAGCCAAAATCTTAAATTTCTTCCTTTGTGATGA 501
Qy 472 CCACCGGAGCTCTAGATTAAACAAACAGTCAGCTAGTCATAGAA 516
Db 502 CCGACTGGAGCTTGGATTATCAGACGGGAAGCAGGTTTGGAA 546

RESULT 4
US-10-158-844-36
Sequence 36, Application US/10158844
Publication No. US20040029118A1
GENERAL INFORMATION:
APPLICANT: Kunsch et al.
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-R
COMPUTER: Dell Latitude Pentium 3
OPERATING SYSTEM: Windows 98
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICANT: Kunsch et al.
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/158,844
PRIOR FILING DATE: 03-Jun-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/961,527
FILING DATE: 1997-10-30
APPLICATION NUMBER: US 60/029,960

FILING DATE: 1996-10-31
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Mark J.
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB340P1D1
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 21706 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-10-158-844-36

Query Match 12.4%; Score 75.4; DB 13; Length 21706;
Best Local Similarity 50.4%; Pred. No. 4.5e-13;
Matches 212; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

Qy 99 TGGAGCGTCCGGTTCAGGAAAATCCACTTTGCTCAACTGTCTTGGCACACTTGACAAACC 158
Db 14961 TGGTGCTTCAGGTGCGAGCAAGTCAACAGTCTTAACTTCTTGGGGAATGGATACCAA 15020
Qy 159 AAGTTCGGGACAGATCTTGTGCGAGGATGAGACCTTCTGAAACTCTCTACGCGTAAGCA 218
Db 15021 TGATGAAGGGGAAATCTGGAATGATGGTGTAAATATTCGGGATTAAGTTCACACCGG 15080
Qy 219 ACGGTTATACAGAAAATACGGTGGGCTATTATTTCAGAGATTATGCTTGAATTCCTGAA 278
Db 15081 CACCAATTACCGTAGAATGATGCGGGTTCGTTTTCAGTTTATAATCTAGTTCTTAA 15140
Qy 279 CAGGACAGTTAAATTCACCTTCAGCTTGGGTGGGTGGAACACAAATGGCTGAAATTC 338
Db 15141 TCTGACAGTAAAGGAAAATGGAACCTGCTTCTGAAATTTGTGACAGATGCTTGAATTC 15200
Qy 339 ---TCAAGTACTTTCATGCTGTTGTTGCTTGGTTCGAGTCGTCGAGGAAAAGCCAGTTT 395
Db 15201 TGATCAGGCTTCAGAGATGATGCTGCTCATCGTCTCAATACTTCCAGCCAGCT 15260
Qy 396 CTCTGGTGGGCAACAAACAGACTGCGTGGCCCGGCTACTGCTCAAAAATTCCTGAAAT 455
Db 15261 TTCTGGAGGGAGCAACAGCAGTCTCCATTGSCACGCGGTAGCCAAAATTCCTAAAAT 15320
Qy 456 AATTCTGGCTGATGACCAACCGGAGCTCTAGATTTTACAAACAGCTGAGCTAGTCATAGA 515
Db 15321 TCTCCTTTGTGATGAACCGACTGGAGCTTGGATTATCAGAGGGCAAGCAGGTTTGA 15380
Qy 516 A 516
Db 15381 A 15381

RESULT 5
US-10-398-221-3886/c
Sequence 3886, Application US/10398221
Publication No. US20040018514A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederik
ADDRESSEE: GLASER, Philippe
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
PRIOR FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3886
LENGTH: 12278
TYPE: DNA
ORGANISM: Listeria monocytogenes 4b
FEATURE:

```
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3886

Query Match      12.3%; Score 75; DB 16; Length 12278;
Best Local Similarity 46.5%; Pred. No. 4.4e-13;
Matches 280; Conservative 0; Mismatches 316; Indels 6; Gaps 1;

QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTGGCGTTTCGATCTTATGGCAAGGTCCTC 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9490 ATGATGATAGTAAATGTCAGTAAAGATACAGATAAATGATTTTGGAAAGATC 9431
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 AGTCATAAGTTTACCAAGGAAATGACAGACTGCTGAGCGTCGGTTTCAGGAAA 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9430 TCTCTTTCTATTGGAGCTGGTGAATTTATTGAGTCGTCGCGAGAGTGGTGGGAAG 9371
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 TCGACTTTGCTCAACTGCTTGGCACACTTCACAAACCAAGTTCCGGACAGATCCTTGTG 180
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9370 ACCAGCTGCTAAATGTTATCGACACCTAGATTCAAGAATAGTGGAAAGTTATT 9311
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GAGGATGATAGCTTCTGAAACTCTCTACGGTAAAGCAACGGTTATACAGGAAATAACG 240
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9310 AACGAGATGGAATATCAGA-----CGAAAAAAGAGGTTATGACTCTAAAAAAGAGG 9257
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GTGGCTATTTATTTCAAGATTATGCTTGTATCCCGACAGGACAGTAAATTTCAACCTT 300
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9256 TTAGTTTTTATTTCCAAAATTTCTATTGATGGAATGAACAGTGTAGAAAACCTTA 9197
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 CAGCTTGGCGTGGAAAAACAAATATGCGCTGAAATTCCTCAAGTACTTCTGCTGTTGGT 360
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9196 TCCATTACAGGTGGAAAAATCGCAAGCTGATGATAGAGCATTTAGAAAGTAGGAATG 9137
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 CTGAGTCGTCGAGGAAAGCCAGTTTGTGAATCTCTGTCGCGAACACACAGCACT 420
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9136 GATGAAGCTATTTAGCAAAAAAAGTATACCAATTAAGCGGTGGAGAAAAACAAACGGATT 9077
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 GGGTTGGCGCGGTACTGCTCAAAAAATCCCGGAATATTTCTGGCTGATGAACCAACCGGA 480
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9076 GCTATTGTGGCAATTTTACTCAACCAATTTTCAGTTTACTTTCGGACGAACCAACAGGC 9017
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 GCTCTAGATTTAAACAAAGTACGCTAGTCTATAGAGCATTTGACAGCACTGCGCGCAAA 540
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9016 AACTTGGATGATAAAACAAACAAATAATCATTTAGATTTATTTTAGCATTTGAAAAACAA 8957
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 GCGCCACCGTTGTTGTTGCTACGCACTCGCCCTCTTCCGAGAAATCAGCGGATACCAATT 600
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 8956 GGTAAACAAATCAATTTGTGTACATGACCCAGAAATATCTGAAAGCAGATCGAATC 8897
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 AT 602
DB |||||
QY 8896 AT 8895
DB |||||

RESULT 6
US-10-282-122A-27244
; Sequence 27244, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27244
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
; US-10-282-122A-27244

Query Match      12.2%; Score 74.6; DB 13; Length 1401;
Best Local Similarity 50.1%; Pred. No. 1.7e-13;
Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;

QY 100 GGAGCGTCGGTTCAGGAAATCGACTTTGCTCAACTCTCTGAACTCTCTACGCGTAAGCAA 159
DB 805 GGCCTTCAGGATCTGGTAAACCAACCATTAATAACATATTTTCAGGAGATGATAGACT 864
QY 160 AGTCCGGACAGATCCTTGTGAGGATGTAGACCTTCTGAACTCTCTACGCGTAAGCAA 219
DB 865 TCTAGTGTAGTGTATTGTCTCAATGGTTATACATGATTGTGTTAAATGATAGAAAGCTC 924
QY 220 CGGTATACAGAAAAATACGGTGGGCTATTATTTTCAAGATTATGCTTGAATCCCGAC 279
DB 925 ACTAAATTCGCTCAAAAGTATGTTGGTTACATCTTTCAACAATACGGTTTATTACCTAAT 984
QY 280 AGGACAGTTAAATTCACCTTCAGCTTCGGTGGGAAAAACAC-----AAATGG 327
DB 985 TTAAACAGTTAGAAAAACATTTGAGATAGAGCAATCTTCAACAGATCTTAGTAAAGG 1044
QY 328 CCGTAAATTCCTCAAGTACTTCACTGCTGTTGCTGTTAGTCTGTCGAGAAAAAGCCAGTT 387
DB 1045 ATCAGCATTTGATGCACCTTTTAGAGCGGTGGGATGATGTTTGCAAAAGAGCTTCCT 1104
QY 388 TTTGAACCTCTGGTGGCGACACAAAGCACTGGTTGGCCGGTACTCTCAAAAT 447
DB 1105 AATGAATGAGTGGTGGGCAACAGCAACGCTGTTTCCATTGCAAGAGCTTTTGTATAAAC 1164
QY 448 CCGCAATTAATTCGCTGATGAACCAACCGAGCTCTAGATTTTAAACAAACAGTGAGCTA 507
DB 1165 CCTTATTAATTTTGGTGTAGTGAACCTACTGCGGCACCTTGATCTTGAGATGACCCAAAT 1224
QY 508 GTCATAGAAAGCATTTGAGAGCACT 530
DB 1225 GTTTAAACACAGTTTTTTAGCAAT 1247

RESULT 7
US-09-960-870-13
; Sequence 13, Application US/09960870
; Publication No. US20030134281A1
; GENERAL INFORMATION:
```

; APPLICANT: Evans, Glen
 ; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
 ; FILE REFERENCE: P-EA 4738
 ; CURRENT APPLICATION NUMBER: US/09/960,870
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 31241
 ; TYPE: DNA
 ; ORGANISM: M. genitalium
 ; US-09-960-870-13

Query Match 12.2%; Score 74.6; DB 10; Length 31241;
 Best Local Similarity 50.1%; Pred. No. 1e-12;
 Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;

QY	100	GGAGCGTCCGGTTCAGGAAATCGACTTCTGCTCAACTGCTTGGCAGACCTTGACAAACCA	159
DB	8214	GGCCCTTCAGGATCTGGTAAACCCACATTACTTAAACATTTATTCAGGATGGATAGAGCT	8273
QY	160	AGTTCCGGACAGATCCCTTGTGAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCAA	219
DB	8274	TCTAGTGGTAGTGTATTGTCAATGGTTATTAACATGATTTGTTAAATGATAGAAAGCTC	8333
QY	220	CGTTTATACAGGAAATAATACGGTGGGTATTTTCAAGATTATGCTTGAATCCCGAC	279
DB	8334	ACTAAATTCGGTCAAAAGTATGTTGTTACATCTTTCAACAATACGGTATTATACCTAAT	8393
QY	280	AGGACAGTTAAATTCACCTTCAGCTTCGGTGGGAAACAC	327
DB	8394	TTAACAGTTAGAGAAACATTTGATAGAGGAGCAATCTTCAACAGATCCTAGTAAAGG	8453
QY	328	CCTCAATTCCTCAAGTACTTTCATGCTGTTGGTCTTGAAGTGGTTCGAGGAAACCCAGTT	387
DB	8454	ATCAGCATTTGATGACATTTTAGAGCGTTGGGATGGATAGTTTGGCAAGAAGCTTCT	8513
QY	388	TTTGAACCTCTCTGGTGGGAAACAAAGAACTGCGTTGGCCCGGTACTGCTCAAAAT	447
DB	8514	AATGAATTCAGTGGTGGGCAACAGCAACGTTTCCATTGCAAGAGCTTTTGCTTAAAC	8573
QY	448	CCCCGAATAATTCCTGGCTGATGAACCAACCGAGCTCTAGATTAAACAAACAGTGAGCTA	507
DB	8574	CCCTTATTAATTTTGGTGAAGAACTTCTGGGCACTTGTATCTTGATGATGACCCAAAT	8633
QY	508	GTATAGAGCAATTTGAGAGCACT	530
DB	8634	GTTTTAAACAGTTTTTAGCAAT	8656

RESULT 9

US-10-251-668-13
 ; Sequence 13, Application US/10251668
 ; Publication No. US20040063097A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Evans, Glen
 ; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
 ; FILE REFERENCE: P-EA 5441
 ; CURRENT APPLICATION NUMBER: US/10/251,668
 ; CURRENT FILING DATE: 2002-09-20
 ; PRIOR APPLICATION NUMBER: US 09/960,607
 ; PRIOR FILING DATE: 2001-09-20
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 31241
 ; TYPE: DNA
 ; ORGANISM: M. genitalium
 ; US-10-251-668-13

Query Match 12.2%; Score 74.6; DB 13; Length 31241;
 Best Local Similarity 50.1%; Pred. No. 1e-12;
 Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;

QY	100	GGAGCGTCCGGTTCAGGAAATCGACTTGTCTCAACTGCTTGGCAGACCTTGACAAACCA	159
DB	8214	GGCCCTTCAGGATCTGGTAAACCCACATTACTTAAACATTTATTCAGGATGGATAGAGCT	8273
QY	160	AGTTCCGGACAGATCCTTGTGAGGATGTAGACCTTCTGAAACTCTCTACGCGTAAGCAA	219
DB	8274	TCTAGTGGTAGTGTATTGTCAATGGTTATTAACATGATTTGTTAAATGATAGAAAGCTC	8333
QY	220	CGTTTATACAGGAAATAATACGGTGGGTATTTTCAAGATTATGCTTGAATCCCGAC	279
DB	8334	ACTAAATTCGGTCAAAAGTATGTTGTTACATCTTTCAACAATACGGTATTATACCTAAT	8393
QY	280	AGGACAGTTAAATTCACCTTCAGCTTCGGTGGGAAACAC	327

RESULT 8

US-09-960-858-13
 ; Sequence 13, Application US/09960858
 ; Publication No. US2003013877A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Evans, Glen
 ; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
 ; FILE REFERENCE: P-EA 4974
 ; CURRENT APPLICATION NUMBER: US/09/960,858
 ; CURRENT FILING DATE: 2001-09-20
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 31241
 ; TYPE: DNA
 ; ORGANISM: M. genitalium
 ; US-09-960-858-13

Query Match 12.2%; Score 74.6; DB 10; Length 31241;
 Best Local Similarity 50.1%; Pred. No. 1e-12;
 Matches 222; Conservative 0; Mismatches 209; Indels 12; Gaps 1;

Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 34668
LENGTH: 756
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-10-282-122A-34668

Query Match 12.2%; Score 74.4; DB 13; Length 756;
Best Local Similarity 50.0%; Pred. No. 1.4e-13;
Matches 229; Conservative 0; Mismatches 211; Indels 18; Gaps 1;

Qy 57 TCTCAGTCATAAGTTTACAGAACAAATGACAGCACTGACCTGGAGCGTCCGGTTTCAGG 116
Db 72 TATAAGTTTTCAGTTCGAAAGGGGAATTTGTAGCTGTCTATGGTCCCTTCTGGATCAGG 131

Qy 117 AAAATCGACTTTGCTCAACTGTCTTGGCACACTTGACAAACCAAGTTTCGGACAGATCCT 176
Db 132 TAAACGACITTACTCAATGTGTTAAGTTCAATAGACTATTTCAGGAGAACTGTAGA 191

Qy 177 TGTGAGAGATGACACCTCTGAACTCTCTACGCTAAGCAACGGTTATACAGGAAAA 236
Db 192 AGTTGAAGGTAAAGAAATTAATAAATGAGCCACAAAGAGTGGCAAAATTTTCGAAACA 251

Qy 237 TAAGGTGGCTTATTTATTTCAAGATTATGCTTGAATCCCGACAGACAGTAAATTCAA 296
Db 252 ACATCTCGGTTTATTTTTCNAGATTATAGGTTTATACCATTAACAGTAAAGAAA 311

Qy 297 CCTTCAGCTTCGGTGGAAAAACAATAATGCTTGAATTCCTC----- 340
Db 312 TATATGCTACCACTCTCAGTACAAAAATTCATAATATCAAAATGGAACAAAAATTATA 371

Qy 341 --AGTACTTCATGCTCTGGTCTTGTAGCTGTCGAGGAAGGCCAGTTTTCGACTCTC 398
Db 341

Db 372 AGAGTGGCTGAGGCATTAGGTATTTATTAACCTGGGAAATAAATATATCAAGTGAATTC 431
Qy 399 TGGTGGCGAACAAACAAACGAACTGCTTGGCCCGGCTACTGCTCAAAAATCCCGCAATAAT 458
Db 432 TGGCGGTGAGCAACAAACGTAACGGGGGAGCCCGGGCATTCGTCATAAACCAACAATTAT 491
Qy 459 TCTGGCTGATGAACCAACACCGGAGCTCTAGATTAAACAA 496
Db 492 TTTGCGAGATGAACCTACTGCGCATTAGATTCTAAAA 529

RESULT 13
US-10-398-221-3820
Sequence 3820, Application US/10398221
Publication No. US20040018514A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederik
APPLICANT: GLASER, Philippe
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3820
LENGTH: 5312
TYPE: DNA
ORGANISM: Listeria monocytogenes 4b
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(end)
OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3820

Query Match 12.1%; Score 73.6; DB 16; Length 5312;
Best Local Similarity 50.2%; Pred. No. 7.9e-13;
Matches 234; Conservative 0; Mismatches 223; Indels 3; Gaps 2;

Qy 100 GGAGCGTCCGGTTTCAGGAAATCGACTTTGTCTCACTCTTGGCACACTTCTACCGTAAGCAA 159
Db 3955 GGACCTTCTGGAGCGGAAATCCACTGTACTTAATATACITGTGGTAATGATAGTCA 4014

Qy 160 AGTCCGGACAGATCCTTGTGAGATGTAGACCTTCTGAACCTCTCTACCGTAAGCAA 219
Db 4015 AGTGAAGGCAAAATTTATGCTAGATGGCAGGATTTGCTCAATATAATGCAAAACAATA 4074

Qy 220 CGTTTATACAGGAAATATACGCTGGCTTATTTTCAAGATTATGCTTGAATCCCGAC 279
Db 4075 ACCAATATCGCGNACAGATGTGGCTTGTATTTCAATTTTACATTTGTACCCCAAT 4134

Qy 280 AGGACAGTTAAATTAACCTTCAGCTTCGGTGGGAA-ACACAAATGGCTGAAATTC 338
Db 4135 TTGACTGTCTAAAGAGACGTAGAAATTAGTCTGCCAAATTTGGCCAAATGCAATTAGACGA 4194

Qy 339 TCAAGTACTTCA--TGCTGTTGGTCTTTCAGTCTGAGGAGAAAGCCAGTTTTCGAACTC 396
Db 4195 GAACTGTTTTAACTCAAGTTGGATTAAAGTACAGATTAGACAATTTTCAGCACAATA 4254

Qy 397 TCTGTTGGCGAACCAACAGCACTGCTTGGCCGGGTACTGCTCAAAAATCCCGAATA 456
Db 4255 TCTGGCGGGAACCAACAGCGTGTGCCATAGCAGCTGGCTTGTCTTAAAGCACCAGAACTG 4314

Qy 457 ATTCTGGCTGATGAACCAACCGGAGCTCTAGATTTAACAACAGTACGCTAGTATGAA 516
Db 4315 CTTTTATCGATGAGCCAAACCGGGCACTTGATTACGATCTGGAATAATCAGTCTTAAAA 4374

Qy 517 GCATTGAGAGCACTGCCGCAACAAAGGCCCGCCAGCTTGTGTGTGCTA 562
Db 4375 TTATTACAAGAACTTGCCTTAACACGGGAACAACCGTTATTGTTA 4420

RESULT 14
US-10-282-122A-30817
; Sequence 30817, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 30817
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Pasteurella multocida
US-10-282-122A-30817
Query Match 12.0%; Score 73.2; DB 13; Length 684;
Best Local Similarity 46.4%; Pred. No. 3.3e-13;
Matches 283; Conservative 0; Mismatches 318; Indels 9; Gaps 1;
QY 2 TCATAGAAATCAATGACCTCAAGAAATCTTTTGGCGTTGCGATCTTATGGCAAGGTCTCA 61
DB 2 TGATGAAACACACATTTATCAAAAGATTCGTCAGTGACCGCACTTGAAGACATCA 61
QY 62 GTCATAAGTTTTTACCAGGAACAATGACAGCACTGACGTGGAGCGTCGGTTTCAGGAAAT 121
DB 62 ACATTCAAAATTAAGAGAGCGCAATTTGTTGCAATCATGGGTGCTCAGGTTTCAGGCAAAA 121
QY 122 CGACTTTCCTCAACTGCTTTCGCACACTTGCAAAACCAAGTTCGGGACAGATCTCTTCG 181
DB 122 CGACGTTAATGAATTTCTCACTGGTTTAGCACTGCGATGAAGGAAAAGTATCTTAG 181
QY 182 AGGATGTAGACCTTCTGAAACTCTTACGGGTAAAGCAACGGTTTATACAGGAAATAACGG 241
DB 182 ATGGTGTGATGCGGCACAACTTGAATGAATTCGGCAGACACACGATTCGGGCGGAGAGA 241
QY 242 TGGGCTATTATTTCAGATATTGCCTTGATTCCCGACAGGACAGTTAAATTCACACCTTC 301

DB 242 TCGGTCTGGTCTTTTCAGCAATTCACCTTAATCCCTTATTTAACTGCCTTGGAACGTC 301
QY 302 AGCTTGGCGTGAACAAACACAAATGGCCTGAAATTCCTCAA-----GTACTTCTATG 352
DB 302 TGTAGCACACATATATCACAGTGAATGATGAAGATGCGAGCCGAGCGGTGTATCAAC 361
QY 353 CTGTTGGTCTTCAGTTCGTTTCAGAGGAAAAAGCCAGTTTGTGAACCTCTCTGTTGGCGAAACAA 412
DB 362 AAGTCGGTCTGGCGCATCGTTTCGACACAGACCAAGCAACTGTCTGCGGAGACAA 421
QY 413 AAGCAACTGCGTTGGCGCGGTACTGTCAAAAATCCCGAATTAATTCGTGCTGATGAC 472
DB 422 AAGCTGTGTGTTATTCGCGCGCTTAGTTAATCAGCCGCGGTGATTTTCCCGATGAGC 481
QY 473 CAACCGGAGCTCTAGATTTAAACAAACAGTGTAGTCTATAGAACATTTGAGAGCACTCG 532
DB 482 CGACAGGAAACCTAGATGAAAAAANTGAGCACTGCTGCTCGATTACTCGTCACATTAA 541
QY 533 CGGACAAAGGCCCAACCGTTGTTGTTGTTGCTAGCAGCTCGCCCTCTTCGAGAAATCAGCG 592
DB 542 ACCAAACAGGACGTACAGTGTGATGTCACCCCAATATCTGAGCTAAGTAACTGCGCG 601
QY 593 ATACCATTAAT 602
DB 602 ACCGCACTAT 611
RESULT 15
US-09-738-626-956
; Sequence 956, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 956
; LENGTH: 666
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-956
Query Match 11.9%; Score 72.4; DB 9; Length 666;
Best Local Similarity 47.4%; Pred. No. 6e-13;
Matches 257; Conservative 0; Mismatches 276; Indels 9; Gaps 1;
QY 76 CCAGGAACAATGACAGCACTGACTGGAGCGTCCGGTTTCAGGAAATCAGCTTTGCTCAAC 135
DB 106 CCGCGGAACCTGCTGTCATCATGGCCCGTCCGGCTCAGGAAATCCACCTTCTCAAC 165
QY 136 TGTCTTGGCACATTTGACAAACCAAGTTCCGGACAGATTCCTTGTGAGGATGTAGACCTT 195
DB 166 GTCCGCGGCTCTCTGACGGGCAACCTCTGCGCATGTGCTTATCGACGTCGCCGCGCC 225

QY	196	CTGAACCTCTCTACGGTAAAGCAACGGTTATACAGGAAAAATACGGTGGCTATTATT	255
Db	226	TCAGACCTCAACCGCCAAACGGCAGCTGAACACCGCGCTCGGCACATCGAGTTATTTTC	285
QY	256	CAAGATTATGCTTGTATCCCGACAGGACAGTTAAATTCACCTTCAGCTTCGGTGGAA	315
Db	286	CNAACCTACAACTGTGTCCCAACCTCACCGTCGGAGAAAACGTGGTCTGCCCTTAGAA	345
QY	316	-----AAACACAATGGCTGAAATTCCTCAAGTACTTCATGCTGTTGCTTGGAG	366
Db	346	CTCGACGGCAAAACCGACCGCCAGGAGTAGCAATCGCACTCGGGAAGTCGGCCCTCGAA	405
QY	367	TCGTTGAGGAAAAAGCCAGTTTTTGAACCTCTCTGGTGGCAACACACGAACTGCGTTG	426
Db	406	GGCTTCGACGACCGCTTCCCGGAGAGATCTCGGCGGCCAAGCCAGCGCTCGCCATC	465
QY	427	GCCCGGTACTGCTCAAAATCCCGAATAATTCCTGGCTGATGAACCAACCGGAGCTCTA	486
Db	466	GCGAGAGCCCTCATCGGCCCGCGAAAAATCTTGCTTCCGACGAAACCCACCGCGCCCTC	525
QY	487	GATTTAAACAAACAGTGAAGCTAGTAGAGCATTCAGAGCACTCGCCGACAAAGCGGCC	546
Db	526	GACACCTCCACCGGACGCACTCTCCGGCTCTCCGCCAAAGAATCGATTCCGGTGCC	585
QY	547	ACCGTTGTTGTTGCTACGCACTCGCCCTCTTCCGAGAAATCAGCGGATACCATTTATCAA	606
Db	586	GCAGGCTCTCTGTCAACCAAGAACCCCGCTTCGCGCGTGGGCGAGACCGAAACAATCATG	645
QY	607	CT 608	
Db	646	CT 647	

Search completed: July 30, 2004, 15:28:19
Job time : 518 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2004, 07:32:14 ; Search time 2796 Seconds
(without alignments)
6504.316 Million cell updates/sec

Title: US-09-868-338-7_COPY_1117_1725

Perfect score: 609

Sequence: 1 atgatgaaatcaatgacct.....cggataccattatcaacta 609

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estma.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hc.*

9: gb_est1.*

10: gb_est2.*

11: gb_hc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pin.*

20: em_gss_vit.*

21: em_gss_fun.*

22: em_gss_man.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75.2	12.3	647	14	CD164788
2	63.4	10.4	622	28	BH861070
3	62.2	10.2	639	29	CG954936
4	62.2	10.2	668	14	CA847456

5	61	10.0	404	14	CD740366
6	61	10.0	542	14	CD731492
7	61	10.0	619	9	AL586817
8	61	10.0	743	13	BU399305
9	61	10.0	798	13	BU328328
10	61	10.0	896	13	BU407162
11	61	10.0	907	13	BU112595
12	60.4	9.9	759	13	BU865169
13	59.4	9.8	576	9	AU090519
14	59.2	9.7	358	14	H54613
15	59.2	9.7	426	10	BF700602
16	59.2	9.7	448	10	BF691964
17	59.2	9.7	481	13	EX281688
18	59.2	9.7	528	9	AA456377
19	59.2	9.7	542	12	BM126834
20	59.2	9.7	584	12	BT195085
21	59.2	9.7	598	9	AT807162
22	59.2	9.7	646	14	CD672146
23	59.2	9.7	667	13	EX509020
24	59.2	9.7	761	13	EX106160
25	59.2	9.7	919	13	EX403383
26	59.2	9.7	923	13	BQ229972
27	59.2	9.7	963	12	BG387740
28	59.2	9.7	1201	9	AL520322
29	58.4	9.6	296	10	AY125863
30	57.8	9.5	713	9	AU213416
31	57.8	9.5	922	12	BG442902
32	57.6	9.5	780	9	AV726232
33	57.6	9.5	880	13	BX392299
34	57.6	9.5	1065	13	BX325244
35	57.6	9.5	4035	11	EC042531
36	57.4	9.4	790	14	CB894287
37	57.2	9.4	300	28	AZ578156
38	57	9.4	833	14	CB894924
39	56.8	9.3	515	28	BZ331622
40	56.8	9.3	566	9	AU165128
41	56.8	9.3	655	28	BZ348974
42	56.8	9.3	1201	9	AL520321
43	56.4	9.3	3843	29	AY408954
44	56.4	9.3	3843	29	AY408955
45	55.4	9.1	780	12	BG587938

ALIGNMENTS

RESULT 1
CD164788
LOCUS
DEFINITION
MLJ-0091T-L258-D10-U.B MLJ-0091 Schistosoma mansoni CDNA Clone
MLJ-0091T-L258-D10.B, mRNA sequence.
ACCESSION
CD164788
VERSION
EST.
KEYWORDS
SOURCE
ORGANISM
Schistosoma mansoni
Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE
1 (bases 1 to 847)
Verjovskii-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,
Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.P.,
Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
Nascimento, A.L.T.O., Ohlweiler, F.F., Reis, E.M., Ribeiro, M.A.,
Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,
Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acelomate human parasite Schistosoma mansoni

TITLE
CD164788 MLJ-0091T
JOURNAL
BH861070 S2H2SP6 W
MEDLINE
CG954936 MBEPIOTF
COMMENT
CA847456 EST0364 C
Contact: Dr. Sergio Verjovskii-Almeida

Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil

Tel.: +55-11-3091-2173

Fax: +55-11-3091-2186

Email: verjoe@iq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL <http://bioinfo.iq.usp.br/schisto/>

Plate: MLI-00917-L258 row: 10 column: D.

FEATURES

Location/Qualifiers
1..647
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MLI-00917-L258-D10.B"
/sex="mixed pool"
/dev_stage="miracidium"
/clone_lib="MLI-0091"
/note="Vector: pGEM T-easy"

ORIGIN

Query Match 12.3%; Score 75.2; DB 14; Length 647;
Best Local Similarity 53.8%; Pred. No. 2.4e-12;
Matches 155; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
QY 1 ATGATAGAAATCAATGACCTCAAGAAATCTTTTGGCGTGTGGATCTTATGGCAAGGCTC 60
DB 222 ATGCTGAAATCTCTGAGCTGGCCAAATCGTTTGGCTCCAGAGCGTTGTGGAGGGCCG 291
QY 61 AGTCATAAGTTTTACCAAGGAACATGACAGCAGTCTGAGCGTCCGTTTACAGGAAA 120
DB 292 AACTGACGTTTCATGCGCGGAAGGATGACACCGTGTGGTTCGAGTGGCTCGGAAAA 351
QY 121 TGGATTGCTCACTGTCITGGCAGACTTGACAAACCAAGTTCGGACAGATCCTTGTC 180
DB 352 TCGACACTGTTGAATGCTTGGAAACATTTGATCGTCCAAAGGTTGAGCGTATCCTTTGG 411
QY 181 GAGGATGTAGACCTTCTGAAACTCTCTACGGTGAAGCAACGGTTTATACAGAAAAATACG 240
DB 412 AATGTAAGATGTGCTGCTGCAAGCGTCTGTCAGAGGCGACGATGCGGAAATCTGAC 471
QY 241 GTGGCTATTTATTTCAAGATATGCTTGTGATTCGGACAGGACAGTT 288
DB 472 CTCGGTTATTTATTCAGAACTACGCAATGTTGTTGAGCAGCCCGTT 519

RESULT 2

BH861070 622 bp DNA linear GSS 29-MAY-2003
LOCUS S2H2SP6 Western x-disease phytoplasma SpeI cosmid library Western x
phytoplasma genomic clone S2H2 similar to ABC transporter, ATP
binding protein (pHnL), genomic survey sequence.

ACCESSION

VERSION BH861070.1 GI:31126928

KEYWORDS

SOURCE Western X phytoplasma

ORGANISM

Western X phytoplasma
Bacteria; Firmicutes; Mollicutes; Acholoplasmatales;
Acholoplasmataceae; Phytoplasma; 16SrIII (X-disease group).

REFERENCE

AUTHORS Liefing, L.W. and Kirkpatrick, B.C.
TITLE Cosmid cloning and sample sequencing of the genome of the
uncultivable mollicute, Western X-disease phytoplasma, using DNA
purified by pulsed-field gel electrophoresis
JOURNAL FEMS Microbiol. Lett. 221 (2), 203-211 (2003)
MEDLINE 22612896
PUBMED 12725928

COMMENT

Contact: Liefing LW
Department of Plant Pathology
University of California, Davis

One Shields Ave, Davis, CA 95616, USA

Tel: 530 752 1697

Fax: 530 752 5674

Email: cwlieting@ucdavis.edu

Sequences were manually edited using Sequencher to remove vector
and low quality sequence.

Seq primer: SP6

Class: cosmid ends.

Location/Qualifiers

1..622

/organism="Western X phytoplasma"

/mol_type="genomic DNA"

/strain="Jensen"

/db_xref="taxon:37704"

/clone_lib="Western X-disease phytoplasma SpeI cosmid
library"

/note="Vector: pcosRW2; Site 1: SpeI; Pulsed-field gel
purified DNA from Western X-disease phytoplasma was
digested to completion with SpeI and cloned into the AvrII
site of pcosRW2. pcosRW2 was first modified by inserting
an annealed oligonucleotide containing the recognition
sequences for AvrII, NcoI and MluI into the EcoRI site."

ORIGIN

Query Match 10.4%; Score 63.4; DB 28; Length 622;
Best Local Similarity 48.5%; Pred. No. 1.4e-08;
Matches 215; Conservative 0; Mismatches 216; Indels 12; Gaps 1;
QY 100 GGAGCGTCCGGTTACAGAAATCGACTTTGCTCAACTGCTTGGCACACTTGACAAACCA 159
DB 156 GGAATATCAGATCGGCAATCGACTCTTTAAACTTGTGGAGGACTCGATAAGTGT 215
QY 160 AGTTCGGACAGATCCTTGTGAGGATGTAGACCTTCTGAACCTCTACCGTAAGCAA 219
DB 216 GATAGTGGAGATATATTAAATTAAGTTTCTATAAAGAAATTAACAACAAATGATTA 275
QY 220 CGGTTATACAGAAAAATACGGTGGGCTATTATTTCAGAGATTATGCTTGTGATTC 279
DB 276 GACGACTATCGAAGCGGCTTTAGCTTTATTTCGAAGAATTTAATTTAATAGAAAT 335
QY 280 AGGACAGTTAAATTCACCT-----TCAGTTTCGGTGGAAAAACACAAATGG 327
DB 336 ATCAATGTTTATGAAATATCGCTTTAGCTTAATGAGCTTCAGGGCCAAAAACCGACA 395
QY 328 CCTGAAATTCCTCAAGTACTTCATCTGTTGCTTGTAGTCTGTCGAGGAAAAAGCCAG 387
DB 396 AAACATATTATTAACCTTATTAAGAANAATGGAACCTTTAGTTATGAAGAAGAAAA 455
QY 388 TTTGAACCTCTGTTGGGGAACAAACAGAACTGGCTTGGCCCGGTTACTGCTCAAAAT 447
DB 456 AACGAATATCAGCGGCAAAAAACAAAGAAATAGCCATCGCTCGTGTCTTTGGTTAA 515
QY 448 CCCCAGATTAATTCGCTGATGAACCAACCGAGCTCTAGATTTTAAACAAACAGTGAGCT 507
DB 516 CCGGAGATTATTTTACGGATGAACCAACAGGAACTTGATTCAAAACCTGGCGAACAA 575
QY 508 GTCATAGAGACATTGAGAGCACT 530
DB 576 ATTTTTAAATTTTAAAAAAT 598

RESULT 3

CG954936

LOCUS MBEP107F mth2 Medicago truncatula genomic clone 20B20, genomic

survey sequence.

ACCESSION CG954936

VERSION CG954936.1 GI:39870005

KEYWORDS GSS.

SOURCE Medicago truncatula (barrel medic)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

CG954936 639 bp DNA linear GSS 15-DEC-2003
MBEP107F mth2 Medicago truncatula genomic clone 20B20, genomic
survey sequence.

ACCESSION CG954936

VERSION CG954936.1 GI:39870005

KEYWORDS GSS.

SOURCE Medicago truncatula

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE 1 (bases 1 to 639)
 AUTHORS Town.C.D., Shetty,J., Koo,H. and Feldblyum,T.P.
 TITLE Sequencing of BAC ends from Medicago truncatula
 JOURNAL Unpublished (2003)
 COMMENT Other GSSs: MBEBP10TR
 Contact: Chris Town

TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523
 Fax: 301-838-0208

Email: cdtown@tigr.org

Seq primer: TGTAAACGACGGCCAGT

Class: BAC ends.

FEATURES

source
 Location/Qualifiers
 1..639
 /organism="Medicago truncatula"
 /mol_type="genomic DNA"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone_lib="20B20"
 /note="Vector: pBeloBAC11; Site 1: HindIII; Site 2: HindIII; Cook, D.R. and Kim, D.J., unpublished"

ORIGIN

Query Match 10.2%; Score 62.2; DB 29; Length 639;
 Best Local Similarity 56.7%; Pred. No. 3.3e-08;
 Matches 115; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 358 GGTCTTGAGTCGTCGAGGAAAGCCAGTTTGAACCTCTCTGTCGGCGGACCAACGA 417
 |||||
 DB 87 GGTCTTGAGTCGTCGAGGAAAGCCAGTTTGAACCTCTCTGTCGGCGGACCAACGA 416
 |||||

QY 418 ACTGCGTTCGTCGAGGAAAGCCAGTTTGAACCTCTCTGTCGGCGGACCAACGA 477
 |||||
 DB 147 ATTGCATTCGAGGGCCATTCGAGATCCAGATCTCTTCTTGATGAGCAACA 206
 |||||

QY 478 GGAGCTCTAGATTAAACAAGTCGAGCTAGTCATGAGCAATGAGAGCACTGCGCGAC 537
 |||||
 DB 207 AGTGCATTCGATGTCGAGTCGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCA 566
 |||||

QY 538 AAAGGCGCCACCGTTGTTGTC 560
 |||||
 DB 267 CAAGAGCACTACCGTAGTTGTC 289
 |||||

RESULT 4

CA847456 668 bp mRNA linear EST 16-DEC-2002
 LOCUS EST0364 CM334 Root cDNA Capsicum annuum cDNA clone RR29-81 similar
 DEFINITION to Arabidopsis thaliana putative ABC transporter protein, mRNA
 sequence.

ACCESSION CA847456
 VERSION CA847456.1 GI:26997386

KEYWORDS

SOURCE Capsicum annuum

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Capsicum.

REFERENCE 1 (bases 1 to 668)
 AUTHORS Richins,R., Alvarado,K., Leger,J. and O'Connell,M.A.

TITLE Responsive transcripts in Phytophthora capsici-challenged roots of
 Capsicum annuum

JOURNAL Unpublished (2002)

CONTACT: O'Connell, MA

Department of Agronomy and Horticulture

New Mexico State University

MSC 30, P.O. Box 30003, Las Cruces, NM 88003, USA

Tel: 505 646 5172

Fax: 505 646 6041
 Email: mocconnel@nmsu.edu

Seq primer: T3

High quality sequence stop: 668.

FEATURES

source
 Location/Qualifiers
 1..668
 /organism="Capsicum annuum"
 /mol_type="mRNA"
 /strains="Criollo de Morelos 334 (CM334)"
 /db_xref="taxon:4072"
 /clone="RR29-81"
 /dev_stage="6 hours after Phytophthora capsici
 inoculation"
 /clone_lib="CM334 Root cDNA"
 /note="Organ: Root; Vector: Zap; Site_1: EcoRI; Site_2:
 Xho 1"

ORIGIN

Query Match 10.2%; Score 62.2; DB 14; Length 668;
 Best Local Similarity 56.7%; Pred. No. 3.4e-08;
 Matches 115; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 358 GGTCTTGAGTCGTCGAGGAAAGCCAGTTTGAACCTCTCTGTCGGCGGACCAACGA 417
 |||||
 DB 162 GGTCTTGAGTCGTCGAGGAAAGCCAGTTTGAACCTCTCTGTCGGCGGACCAACGA 417
 |||||

QY 418 ACTGCGTTCGTCGAGGAAAGCCAGTTTGAACCTCTCTGTCGGCGGACCAACGA 477
 |||||
 DB 222 ATTGCCATCGCAAGAGCAATTTTAAAGATCCAGGATATTACTGTTAGATGAAGCAACA 281
 |||||

QY 478 GGAGCTCTAGATTAAACAAGTCGAGCTAGTCATGAGCAATGAGAGCACTGCGCGAC 537
 |||||
 DB 282 AGTGCATTCGATGTCGAGCTAGTCATGAGCAATGAGAGCACTGCGCGAC 537
 |||||

QY 538 AAAGCGCCACCGTTGTTGTC 560
 |||||
 DB 342 AACAGAACCACTGCTAGTTGC 364
 |||||

RESULT 5

CD740366

LOCUS

DEFINITION

CD740366

VERSION

CD740366.1

GI:32291215

EST.

SOURCE

ORGANISM

Gallus gallus (chicken)

Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 404).

AUTHORS Tassel,C. and Han,J.Y.

TITLE Chicken intestinal lymphocyte EST database as a resource for the

analysis of mucosal immune function

JOURNAL Unpublished (2003)

CONTACT: Hyun S. Lillehoj

Animal Parasite Diseases Laboratory

Animal and Natural Resources Institute, USDA

Bldg.1043, BARC-East, Beltsville, MD 20705, USA

Tel: 3015048771

Fax: 3015045103

Email: hlillehoenri@ars.usda.gov

Single pass sequencing. Bases called and trimmed with phred

0.000925 using options -trim_alt '-trim_fasta. Vector identified

by cross_match using options -minmatch 12 -minscore 12

Plate: 24 row: H column: 07

Seq primer: ATTAGTGACACTATAG

High quality sequence stop: 404.

Location/Qualifiers

1..404

SOURCE

/tissue_type="Gut"		/cell_type="Lymphocyte"		/dev_stage="Adult"		/lab_host="EMDH10B"		/clone_lib="1GAL - Chicken Intestinal Lymphocyte"		/note="Organ: Intestine; Vector: pCMV-SpOrT6; Site 1: SalI; Site 2: NotI; Normalized library from chicken gut infected with coccidia duodenum and middle gut."	
ORIGIN											
Query Match	10.0%;	Score 61;	DB 14;	Length 542;							
Best Local Similarity	60.6%;	Pred. No. 7.5e-08;									
Matches 100;	Conservative 0;	Mismatches 65;	Indels 0;	Gaps 0;							
QY	358	GGCTTTGAGTCGTTGACGAAAAGCCAGTTTTTGGACTCTCTGGTGGCGAACACGCA	417								
Db	270	GGATTTTGACACTGTAGTTTGGAGAAAAGGCATTTACTTTTCAGGTGGACAGAGCAGCA	329								
QY	418	ACTGCGTTGGCCCGGCTACTGCTCAAAATCCCGAATAATTCCTGGCTGATGCAACCAACC	477								
Db	330	ATTGCAATTGCTCGAGCTCTGCTCAGAAATCCCAAAATTCCTGTTAGACGACCAACA	389								
QY	478	GGAGCTCTAGATTTAACAAACAGTCGAGCTAGTCATAGAGCATTTG	532								
Db	390	ATGCTCTCGATGCTGTAATGAGTACCTAGTGCAGGAAGAGCTCTG	434								
RESULT: 7											
AL586817											
LOCUS											
DEFINITION											
ACCESSION											
VERSION											
KEYWORDS											
SOURCE											
ORGANISM											
REFERENCE											
AUTHORS											
TITLE											
JOURNAL											
COMMENT											
FEATURES											
source											
ORIGIN											
Query Match	10.0%;	Score 61;	DB 9;	Length 619;							
Best Local Similarity	60.6%;	Pred. No. 7.9e-08;									
Matches 100;	Conservative 0;	Mismatches 65;	Indels 0;	Gaps 0;							

QY 358 GGTCTTGAGTCGTTGAGGAAAAGCCAGTTTTTTGAACCTCTCTGGTGGGACACAAACGA 417
 |||||
 Db 176 GGAATTTGACACTGTAGTTGGAGAAAAGGCAATTTCTATTTCAGGTGGACAGAGCAGCA 235
 |||||
 QY 418 ACTGCGTTGGCCCGGTACTGCTCAAAAATCCCCGAAATATTCGCTGATGAACCAACC 477
 |||||
 Db 236 ATTGCAATTCGCTGAGCTCTGCTCAGAATCCCAAAATTTCTTGTAGACGAGCAACA 295
 |||||
 QY 478 GGAGCTCTAGATTAAACAACAGTGAAGCTAGTCATAGAAGCATG 522
 |||||
 Db 296 AGTGCTCTGATGCTGAAAATGAGTACCTAGTCAAGAAGCTCTG 340
 |||||

RESULT 8
 BU399305
 LOCUS 603535206F1 CSEQCHN58 743 bp mRNA linear EST 27-NOV-2002
 DEFINITION sequence.
 ACCESSION BU399305
 VERSION BU399305.1 GI:25768361
 KEYWORDS EST
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 743)
 AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..743
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="CHEST49511"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN58"
 /notes="Organ: small intestine; Vector: pBluescript II
 KS(+); Site 1: EcoRI; Site 2: NotI; This normalized
 library was constructed from 1 million independent clones.
 cDNA synthesis was initiated using an oligo(dT) primer,
 using methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN
 Query Match 10.0%; Score 61; DB 13; Length 743;
 Best Local Similarity 60.6%; Pred. No. 8.6e-08;
 Matches 100; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 358 GGTCTTGAGTCGTTGAGGAAAAGCCAGTTTTTTGAACCTCTCTGGTGGGACACAAACGA 417
 |||||
 Db 162 GGAATTTGACACTGTAGTTGGAGAAAAGGCAATTTCTATTTCAGGTGGACAGAGCAGCA 221
 |||||
 QY 418 ACTGCGTTGGCCCGGTACTGCTCAAAAATCCCCGAAATATTCGCTGATGAACCAACC 477
 |||||
 Db 222 ATTGCAATTCGCTGAGCTCTGCTCAGAATCCCAAAATTTCTTGTAGACGAGCAACA 281
 |||||
 QY 478 GGAGCTCTAGATTAAACAACAGTGAAGCTAGTCATAGAAGCATG 522
 |||||
 Db 282 AGTGCTCTGATGCTGAAAATGAGTACCTAGTCAAGAAGCTCTG 326
 |||||

RESULT 9
 BU328328
 LOCUS 603496289F1 CSEQCHN64 798 bp mRNA linear EST 28-NOV-2002
 DEFINITION sequence.
 ACCESSION BU328328
 VERSION BU328328.1 GI:25836329
 KEYWORDS EST
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 798)
 AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..798
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST406a24"
 /tissue_type="whole embryo"
 /dev_stage="10"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN64"
 /notes="Organ: whole embryo; Vector: pBluescript II KS(+);
 Site 1: EcoRI; Site 2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN
 Query Match 10.0%; Score 61; DB 13; Length 798;
 Best Local Similarity 54.1%; Pred. No. 8.9e-08;
 Matches 124; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 358 GGTCTTGAGTCGTTGAGGAAAAGCCAGTTTTTTGAACCTCTCTGGTGGGACACAAACGA 417

Db 333 GGATTGACACTGAGTGGAGAAAAGGCACTTCTTTTCAGGTGGACAGAGCAGCGA 392
 QY 418 ACTCGTTGGCCGGGACTGCTCAAAAATCCCGGAATAATTCGGCTGATGAACCAACC 477
 Db 393 ATTGCATGCTCAGCTGCTGCTCAAGATCCCAAAATTCCTTCTGTAGACGAGCAACA 452
 QY 478 GGAGCTCTAGATTAAACAAAGTACAGTATAGAACATGAGAGACATGCGCGGAC 537
 Db 453 AGTGCTCTGATGCTGAAATAGTACCTAGTGCAGAGAGCTCTGGACCGGCTGATGGCA 512
 QY 538 AAGGGCCCGCTGTTGTTGTACCACTCGCCCTCTTCGAGAAAT 586
 Db 513 AGGAGGACAGCTCTTAATCATGCTACCGTCTGCTTACTATTTCAGAAAT 561

RESULT 10
 BU407162
 LOCUS 896 bp mRNA linear EST 27-NOV-2002
 DEFINITION BU407162.1 CSEQCHN59 Gallus gallus CDNA clone CHEST375f3 5', mRNA sequence.
 ACCESSION BU407162
 VERSION BU407162.1 GI:25776218
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 896)
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..896
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST375f3"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN59"
 /note="Organ: limbs; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research (1996) 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 10.0%; Score 61; DB 13; Length 896;
 Best Local Similarity 60.6%; Pred. No. 9.4e-08;
 Matches 100; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 358 GGTCTTGTAGTCTGTTCGAGAAAAGCCAGTCTTTTGAACCTCTCTGTGGGAAACAACAACGA 417
 Db 162 GGATTGACACTGAGTGGAGAAAAGGCATTCTACTTTTCAGGTGGACAGAGCAGCGA 221
 QY 418 ACTCGTTGGCCGGGACTGCTCAAAAATCCCGGAATAATTCGGCTGATGAACCAACC 477
 Db 222 ATTGCATGCTCAGCTGCTGCTCAAGATCCCAAAATTCCTTCTGTAGACGAGCAACA 281
 QY 478 GGAGCTCTAGATTAAACAAAGTACAGTATAGAACATGAGAGACATG 522
 Db 282 AGTGCTCTGATGCTGAAATAGTACCTAGTGCAGAGAGCTCTG 326

RESULT 11
 BU112595
 LOCUS 907 bp mRNA linear EST 25-NOV-2002
 DEFINITION BU112595.1 CSEQCHL14 Gallus gallus CDNA clone CHEST102h17 5', mRNA sequence.
 ACCESSION BU112595
 VERSION BU112595.1 GI:25316826
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 907)
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..907
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST102h17"
 /dev_stage="22"
 /lab_host="DH10B"
 /clone_lib="CSEQCHL14"
 /note="Organ: head; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; Modification of pBluescript II KS(+) [Stratagene] vector to accommodate cDNA produced with the T-trimmed protocol (Construction of uni-directionally cloned cDNA libraries from messenger RNA for improved 3' end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI. Ligate in double stranded adaptor containing BsgI and BamHI sites [5'ggccgctgcagccgcgcatccgaaaaag] [5'aattcttttttcggatccggggtgcagc]"

ORIGIN

Query Match 10.0%; Score 61; DB 13; Length 907;
 Best Local Similarity 60.6%; Pred. No. 9.4e-08;
 Matches 100; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 358 GGTCTTGTAGTCTGTTCGAGAAAAGCCAGTCTTTTGAACCTCTCTGTGGGAAACAACAACGA 417
 Db 394 GGATTGACACTGAGTGGAGAAAAGGCATTCTACTTTTCAGGTGGACAGAGCAGCGA 453
 QY 418 ACTCGTTGGCCGGGACTGCTCAAAAATCCCGGAATAATTCGGCTGATGAACCAACC 477

Db 468 TGAAGCTCCTCTATTGTTGCTGTGTGATGAAGCAACAAGTGTCTTGGATAGCAACAG 527

Qy 500 GTGAGCTAGTCATAGAAGCATTCAGAGCACTCCGCCACAAAGCGCACCGTTGTTGTTG 559

Db 528 AGCAGAGATACTGACAGCATTGAGGTCACTTGCCCAATRAACGAAACAGCAGTCTTTGTTG 587

Qy 560 CT 561

Db 588 CT 589

RESULT 13

LOCUS AU090519 576 bp mRNA linear EST 02-APR-2002

DEFINITION AU090519 Rice mature leaf Oryza sativa (japonica cultivar-group) cDNA clone S20717, mRNA sequence.

ACCESSION AU090519

VERSION AU090519.1 GI:7608738

KEYWORDS EST.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

AUTHORS 1 (bases 1 to 576)

TITLE Sasaki, T. and Yamamoto, K.

JOURNAL Rice cDNA from mature leaf (2000)

COMMENT Unpublished (2000)

CONTACT: Takuji Sasaki

INSTITUTION National Institute of Agrobiological Resources

ADDRESS Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan

TELEPHONE Tel.: 81-298-38-7441

FAX: 81-298-38-7468

EMAIL: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/PROJECT="RGF".

FEATURES

source Location/Qualifiers

1..576

organism="Oryza sativa (japonica cultivar-group)"

mol_type="mRNA"

cultivar="Nipponbare"

db_xref="taxon:39947"

clone="S20717"

tissue_type="mature leaf"

clone_lib="Rice mature leaf"

ORIGIN

Query Match 9.8%; Score 59.4; DB 9; Length 576;

Best Local Similarity 55.6%; Pred. No. 2.5e-07;

Matches 114; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 357 TGGTCTTGAGTCGTTTGAGAAAAGCGAGTTTGAACCTCTCTGGTGGCGAACAAACAG 416

Db 126 TGGTTATGATCTTTAGTTGCGCCAGCGGTACTCAGCTCTCTGGAGGACAAAACAGAG 185

Qy 417 AACTGCGTTGGCCGGTACTGCTCAAAATCCCGAATAATTCGGCTGATGAACCAAC 476

Db 186 AATTGCAATTGCAAGAGCCATCTTCAAGATCCAAAATCCCTTGGCTCGATGAAGCAAC 245

Qy 477 CGAGGCTCTAGATTAAACAAACAGTGAAGTCTATGAGAGCATTTGAGAGCACTCGCCGA 536

Db 246 AAGTGCACTTGATGTGGAGTCTGAGAGGATAGTTTCAGAGGCACTAATAGATGATGCT 305

Qy 537 CAAGGCGCCACCGTTGTTGCT 561

Db 306 AGAAGAACCACACTCGTTTCGCT 330

RESULT 14

LOCUS H54613 358 bp mRNA linear EST 21-SEP-1995

DEFINITION Y928A09.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:203224 5', similar to SP:S21957 S21957 P-GLYCOPRTEIN PGPI -

```

ARABIDOPSIS ; mRNA sequence.
ACCESSION H54613
VERSION H54613.1 GI:995139
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfs, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
High quality sequence stops: 315
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: M3RP1
High quality sequence stop: 315.
FEATURES
source
Location/Qualifiers
1..358
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3772355"
/db_xref="taxon:9606"
/clone="IMAGE:203224"
/sex="male"
/deb_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen INFLS"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I...oligo(4T) primer
[5' AACGGAAGAATTAATAAGATCTTTTITTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN
Query Match 9.7%; Score 59.2; DB 14; Length 358;
Best Local Similarity 58.5%; Pred. No. 2.3e-07;
Matches 103; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 385 GTTTTGAACCTCTCTGTCGCGACACACGAACTGCGTTCGCGCGGCTACTGCTCAA 444
Db 126 GGTGTTCTCTCTCAGTGGCGACAAACAGCGATTCGCGTTCGCTCTAAAG 185
QY 445 AATCCCGAATAATTCCTGGCTGATGAACCAACCGAGCTCTAGATTAAACAAACAGTGAG 504
Db 186 AATCCCAAAATTCCTCTAGATGAAGCAACAGTCGCGTTCGCGGAAATGAGTAC 245
QY 505 CTAGTCATAGAGCATTCGAGGACCTCGCCGACAAAGCGCCACCGTTGTTGTC 560
Db 246 CTTGTTCAAGAAGCTCTAGATCGACTGATGATGAAGAACGCTGTAGTTATTC 301
Search completed: July 30, 2004, 14:14:26
Job time : 2800 secs

```

```

RESULT 15
LOCUS BF700602 426 bp mRNA linear EST 22-DEC-2000
DEFINITION 502128776f1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4285507 5',
mRNA sequence.
ACCESSION BF700602

```

```

BF700602.1 GI:11986010
EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW1120 row: h column: 20
High quality sequence stop: 426.
Location/Qualifiers
1..426
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4285507"
/tissue_type="Primitive neuroectoderm"
/lab_host="DH10B (Tl phage-resistant)"
/clone_lib="NIH_MGC_56"
/note="Organ: Brain; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGCGCATATGCGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGGCGGACATG-dT(30)BN-3'.
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
ORIGIN
Query Match 9.7%; Score 59.2; DB 10; Length 426;
Best Local Similarity 58.5%; Pred. No. 2.5e-07;
Matches 103; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 385 GTTTTGAACCTCTCTGTCGCGACACACGAACTGCGTTCGCGCGGCTACTGCTCAA 444
Db 238 GGTGTTCTCTCTCAGTGGCGACAAACAGCGATTCGCGTTCGCTCTAAAG 297
QY 445 AATCCCGAATAATTCCTGGCTGATGAACCAACCGAGCTCTAGATTAAACAAACAGTGAG 504
Db 298 AATCCCAAAATTCCTCTAGATGAAGCAACAGTCGCGTTCGCGGAAATGAGTAC 357
QY 505 CTAGTCATAGAGCATTCGAGGACCTCGCCGACAAAGCGCCACCGTTGTTGTC 560
Db 358 CTTGTTCAAGAAGCTCTAGATCGACTGATGATGAAGAACGCTGTAGTTATTC 413
Search completed: July 30, 2004, 14:14:26
Job time : 2800 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:28:59 ; Search time 54 Seconds
(without alignments)
1062.169 Million cell updates/sec

Title: US-09-868-338-9
Perfect score: 1019
Sequence: 1 MEINDLKKSGFVRLWQL.....VVVATHSLPFRSADTIKML 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04: *
1: Genesecp1980s: *
2: Genesecp1990s: *
3: Genesecp2000s: *
4: Genesecp2001s: *
5: Genesecp2002s: *
6: Genesecp2003as: *
7: Genesecp2003bs: *
8: Genesecp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1019	100.0	203	3	AAB12592 Brevibact
2	408	40.0	213	6	ABU01071 S. pneumo
3	399	39.2	213	6	ABU02491 S. pneumo
4	399	39.2	213	6	ABP81472 Streptoco
5	386	37.9	213	6	ABU00443 S. pneumo
6	385	37.8	213	2	AAW80649 S. pneumo
7	380.5	37.3	210	6	ABU02461 S. pneumo
8	380.5	37.3	210	6	ABP81469 Streptoco
9	373	36.6	211	5	ABBS3309 Lactococc
10	362.5	35.6	465	6	ABU36156 Protein e
11	359	35.2	258	4	AAU45701 Propionib
12	359	35.2	258	6	ABM42220 Propionib
13	353	34.6	270	7	ADC97592 E. faeciu
14	353	34.6	466	6	ABU35504 Protein e
15	352	34.5	237	5	ABP26089 Streptoco
16	348	34.2	250	5	ABP26471 Streptoco
17	344.5	33.8	227	6	ABU39077 Protein e
18	344	33.8	207	5	ABB49925 Listeria
19	344	33.8	219	2	AAU34450 Porphyrom
20	344	33.8	238	2	AAU34327 Porphyrom
21	343	33.7	224	5	ABP26472 Streptoco
22	339.5	33.3	222	4	ABP76730 Corynebac
23	339.5	33.3	222	4	AAG90702 C glutami
24	339	33.3	259	5	ABBS5099 Lactococc
25	339	33.3	648	4	AAU38402 Salmonell

26	338.5	33.2	666	5	ABB47488 Listeria
27	337	33.1	252	6	ABU16155 Protein e
28	337	33.1	252	6	ABM72784 Staphyloc
29	336	33.0	236	4	AAU37889 Streptoco
30	336	33.0	249	4	AAU34146 Staphyloc
31	336	33.0	252	4	AAU37113 Staphyloc
32	333.5	32.7	209	2	AAW80648 S. pneumo
33	333.5	32.7	231	4	ABBS2924 Escherich
34	333.5	32.7	231	4	ABBS2924 Escherich
35	333	32.7	233	5	ABB48879 Listeria
36	333	32.7	236	6	ABU02123 S. pneumo
37	332	32.6	252	6	ABU43928 Protein e
38	332	32.6	256	5	ABP39755 Staphyloc
39	331.5	32.5	265	5	ABP65531 Bifidobac
40	331	32.5	234	6	ADB12657 Alloioccc
41	331	32.5	237	6	ADB12655 Alloioccc
42	331	32.5	664	5	ABBS4027 Lactococc
43	330.5	32.4	663	4	AAU36276 Pseudomon
44	330.5	32.4	663	6	ABP59941 Microbial
45	330	32.4	241	6	ABU32430 Protein e

ALIGNMENTS

RESULT 1
AAB12592 standard; protein; 203 AA.
XX AC AAB12592;
XX DT 12-SEP-2003 (revised)
XX DT 09-NOV-2000 (first entry)
XX DE Brevibacterium lactofermentum gltBD ORF2 protein SEQ ID NO:9.
XX KW Brevibacterium lactofermentum; ABC transporter; breeding; ATPase;
XX KW corynebacterium; gltBD; L-glutamic acid.
XX OS Corynebacterium glutamicum.
XX PN WO200037647-A1.
XX PD 29-JUN-2000.
XX PF 16-DEC-1999; 99WO-JP007079.
XX PR 18-DEC-1998; 98JP-00360621.
XX PA (AJIN) AJINOMOTO CO INC.
XX PI Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX XX WPI; 2000-452189/39.
XX DR N-PSDB; AAA65395.
XX AB ABC transporter constituent of Brevibacterium lactofermentum, its encoded gene and variants, applicable in breeding Corynebacteria particularly for production of L-glutamic acid.
XX PS Claim 5; Page 30-31; 34pp; Japanese.
XX CC The present invention describes a protein (I) which can be used to construct an ATP binding cassette (ABC) transporter. ABC transporter DNAs can be used in breeding Corynebacteria particularly for production of L-glutamic acid. The present sequence represents an ORF (open reading frame) protein from the Brevibacterium lactofermentum gltBD gene, which is used in the exemplification of the present invention. (Updated on 12-SEP-2003 to standardise OS field)
XX SQ Sequence 203 AA;
Query Match 100.0%; Score 1019; DB 3; Length 203;

	Best Local Similarity	100.0%;	Pred. No. 1.le-103;			
	Matches 203;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MIEINDLKKSFGVRIILWQGLSHKFLPCTMTALTGASGSKSTLLNCLGTLDKFPSSGQILV	60			
Db	1	MIEINDLKKSFGVRIILWQGLSHKFLPCTMTALTGASGSKSTLLNCLGTLDKFPSSGQILV	60			
QY	61	EDVDLLKLSRKQRIYKNTVGYLFQDYALIPIRTVKENQLQAVEKHKKWPEIPQVLHVG	120			
Db	61	EDVDLLKLSRKQRIYKNTVGYLFQDYALIPIRTVKENQLQAVEKHKKWPEIPQVLHVG	120			
QY	121	LESFEEKVPFELSGEGEQROTALARVLKPNRIILADEPTGALDLTNSELVIEALBALADK	180			
Db	121	LESFEEKVPFELSGEGEQROTALARVLKPNRIILADEPTGALDLTNSELVIEALBALADK	180			
QY	181	GATVVVATHSPILFRESADTIKL	203			
Db	181	GATVVVATHSPILFRESADTIKL	203			

RESULT 2	
ABU01071	
ID	ABU01071 standard; protein; 213 AA.
XX	
AC	ABU01071;
XX	
DT	23-OCT-2003 (revised)
DT	11-FEB-2003 (first entry)
XX	
DE	S. pneumoniae type 4 strain protein from coding region #642.
XX	
KW	Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW	ant inflammatory; antibacterial; immunostimulant; auditory; respiratory;
XX	gene therapy; vaccine.
XX	
OS	Streptococcus pneumoniae; type 4 strain.
XX	
PN	WO200277021-A2.
XX	
PD	03-OCT-2002.
XX	
PF	27-MAR-2002; 2002WO-IB002163.
XX	
PR	27-MAR-2001; 2001GS-00007658.
XX	
PA	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Masignani V, Tettelin H, Fraser C;
XX	
DR	WPI; 2003-040579/03.
DR	N-PSDB; ABX06354.
XX	
PT	New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT	useful as medicaments for treating or preventing a disease or infection
PT	due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT	ear infection.
XX	
PS	Claim 1; SEQ ID NO 1284; 56pp; English.

```

CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying,
CC immunodominant proteins. The present sequence is one of the 269 proteins
CC expressed by the identified coding regions from the genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 213 AA;
Query Match 40.0%; Score 408; DB 6; Length 213;
Best Local Similarity 41.9%; Pred. No. 3e+36;
Matches 88; Conservative 44; Mismatches 68; Indels 10; Gaps 3
QY 1 MIEINDLKSFGVILLWQLGSLHFLPCTMTALTACSGSKGTLLNCLGLTKLPSSQIIV 60
| | | | : | | | : | | | : | | | : | | | : | | | : | | | :
DB 1 MIELKNTSKFEGSQLPSDNLHPHGKKIYALIGTSCGKTLLNNMIGLEPYDKGQIIY 60
| | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 61 EDVDLLKLSTRKQILRYKNTVGYLFDQYALIPIRTVKFNQLAV-----EKHKWEIPISQ 114
| | | | : | | | : | | | : | | | : | | | : | | | : | | | :
DB 61 DGTSLKDI---KPSVFRRDYLYGFQDFGIETSGTVKENLNGLGVKKLKEKISLMKQ 117
| | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 115 VLVHAYGLSEFE-EKPVFELSGGEQORTALARVLKKNPRIILADEPTGALDTNLSLVIA 173
| | | | : | | | : | | | : | | | : | | | : | | | : | | | :
DB 118 ALNRVNLSYLDKQIFELSGGEAQRVALAKIILKPDPPILADEPTASLDPKNSSELLSI 177
| | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 174 LRALADKGATVVVATHSPLFPRESADTIKL 203
| | | | : | | | : | | | : | | | : | | | : | | | : | | | :
DB 178 LESLKQPNRTHIIATHNPLIWEOVDQVIRV 207
| | | | : | | | : | | | : | | | : | | | : | | | : | | | :

```

RESULT 3	
ABU02491	
ID	ABU02491 standard; protein; 213 AA.
XX	
AC	ABU02491;
XX	
DT	23-OCT-2003 (revised)
DT	11-FEB-2003 (first entry)
XX	
DE	S. pneumoniae type 4 strain protein from coding region #2069.
XX	
KW	Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW	ant inflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW	gene therapy; vaccine.
XX	
CS	Streptococcus pneumoniae; type 4 strain.
XX	
PN	WC200277021-A2.
XX	
PD	03-OCT-2002.
XX	
PF	27-MAR-2002; 2002WO-IB002163.
XX	
PR	27-MAR-2001; 2001GB-00007858.
XX	
PA	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Masnani V, Tetelin H, Fraser C;
XX	
DR	WFI; 2003-040579/03.
DR	N-PSDB; ABX07781.

XX New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,
PT useful as medicaments for treating or preventing a disease or infection
PT due to *Streptococcus* bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.

PS Claim 1; SEQ ID NO 4138; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as A8956454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus nucleic acid* sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a *Streptococcus pneumoniae* bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to *Streptococcus* bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences). (Updated on 23-OCT-2003 to standardise OS field)

Sequence 213 AA:

Query Match	39.2%;	Score 399;	DB 6;	Length 213;
Best Local Similarity	42.9%;	Pred. No. 2.9e-35;		
Matches	90;	Conservative 40;	Mismatches 70;	Indels 10; Gaps 3;
Qy	1	MIEINDUKSGFVRIIWQGLSHKFLPGTMTALTGASGSGKSTLNLCLGTLDLKPSSGGQILV	60	
Db	1	MIDIOGLEKKENDRAISGLNKLKLEKGVVALICKSSGKTTLLNLIIGLEKIDGGVLY	60	
Qy	61	EDVLLKLSTRKQRLYRNVTGYLFQDYALIPDRTVFNQLQAVEKKHWEIPOVLHVG	120	
Db	61	QGKDLKTIPTRE---YFRDQGYLFGFQGLLENGSIKENLDLGVGQIKSKVERLERQVG	117	
Qy	121	-LESFE-----EKPVPFELSGGEQORTALARVLKNPRIILADEPTGALLDTNSELVIEA	173	
Db	118	ALEKVNGLYLDLEKIQIYTLSGEAQRVALAKTIILKNPPLILADEPTAALDPENSEEVMNL	177	
Qy	174	LRALADKAGATVVVATHSPLFRESADTIKL	203	
Db	178	LVDLKDENRIIIIIATHNPLVWKNKADEIIDM	207	

RESULT 4
ABP81472
ID ABP81472 standard; protein: 213 AA.

AA ABP81472;

DT 04-MAR-2003 (first entry)

XX DE Streptococcus pneumoniae polypeptide SEO ID NO 389.

Streptococcus pneumoniae: infection: otitis media: antibacterial: KW

XW	diagnosis; gene therapy.	
XX		
OS	Streptococcus pneumoniae.	
XX		
XX		
PN	WO20023855-A2.	
XX		
XX	24-OCT-2002.	
PD		
XX		
XX	12-APR-2002; 2002WO-US011524.	
PF		
XX		
XX	16-APR-2001; 2001US-0283948P.	
PR		
PR	18-APR-2001; 2001US-0284443P.	
XX		
XX		
PA	(AMCY) AMERICAN CYANAMID CO.	
XX		
PI	Zagursky RJ, Masi AW, Green BA, Chakravarti DN, Russell DP;	
PI	Wooters JL;	
XX		
XX	WPI; 2003-093010/08.	
DR		
DR	N-PSDB; AB242320.	
XX		
XX		
PT	New Streptococcus pneumoniae polynucleotides, useful for treating or	
PT	preventing S. pneumoniae infections, or non-systemic diseases, e.g.	
PT	otitis media, which are induced or exacerbated by S. pneumoniae.	
XX		
PS	Claim 42; Page 642-643; 1091pp; English.	

The invention relates to isolated polynucleotides (ABZ72147-ABZ42522) of a *Streptococcus pneumoniae* genomic sequence, a fragment or derivative variant of the polynucleotide or a nucleic acid sequence 95% identical to one of the polynucleotides. The *S. pneumoniae* polynucleotides and encoded polypeptides (ABP81299-ABP91674) are useful for treating or preventing *S. pneumoniae* infections or non-systemic diseases, e.g. otitis media, which are induced or exacerbated by *S. pneumoniae*. These are also useful for detecting *S. pneumoniae* in a biological sample or diagnosing *S.*

CC detecting *S. pneumoniae* in a blood agar sample of a young child.
CC pneumoniae infection in a subject. The polynucleotides have antibacterial
CC activity and are useful in gene therapy

Sequence 213 AA;

Query Match 39.2%; Score 399; DB 6; Length 213;

Query Match	Score	Score Norm	Score Ratio	Score Diff
Best Local Similarity	42.9%	Pred. No. 2.9e-35		
Matches	90; Conservative	40; Mismatches	70; Indels	10; Gaps

Qy 1 MIEINDLKSGFYRILWQGLSHKFLPGTWTALTQASGSGKSTLLNCLGTLDKPSSGQILV 60

DD I MDDI QGSEKFNFKRIFSGEINTENGRVIALI QNGSGKAI I ENITIGRAMEN I DOGRV A I 00

QV 61 EVDLPLK I STRKORLYRNTVGYLFODYALIPDET VKENLQLA VEKHKWPEI POVLHVG 120

Db 61 QGKDLKTTPTRE---YFDQMGIYLFQNGELLENSQIKENLDLGFVGQKTSKVERLERQVG 117

QY	-----EKPVFETSGGROOKTALAKVILKTNPKAILADSF7GAUDDTINSVLVEIA
	-JESFE-
D6	ALEKONTGVIDIEPOTKYTCGSACRVAIAKTIIKNBPILLADPTAAIDPENSEVMN
	: : : :
118	177

QY 174 LRRLADKCATVVVATHSPLFRESADTIKL 203

Db 178 LVPLKDNRIIIITATHNPLVWNKAEIILDM 207

RESULT 5
ABU00443
IN 12000443 04523234 0001054 013 00

XX
AC
ABU00443;

DT	23-OCT-2003	(revised)
DT	11-FEB-2003	(first entry)
...		

DE S. pneumoniae type 4 strain protein from coding region #10.
XX

Db 61 FGOITLAVNSAKSQTIIREHSYLFQNFALIDDETVEYNMLAKVVKPKDKLKKVVE 120
 QY 115 VLHVAVLESFEKVPFELSGEQORTALARVLLKNPRIILADPTGALDITNSLVEIAL 174
 Db 121 ILERVGLSATLHORVLSLGEQORAVARAILKPSQLIILADEPTGSLDPENRDLVKFL 180
 QY 175 RALADKCATVVVATHSPFPRESADTIKL 203
 Db 181 LEWNRGKTVIIVTHDAYVAQQCHRVIEL 209

RESULT 7
 ID ABU02461
 XX ABU02461 standard; protein; 210 AA.
 AC ABU02461;
 XX
 DT 23-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 XX S. pneumoniae type 4 strain protein from coding region #2039.
 XX
 XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
 KW gene therapy; vaccine.
 XX
 XX Streptococcus pneumoniae; type 4 strain.
 XX
 XX WO200277021-A2.
 XX
 XX 03-OCT-2002..
 XX
 XX 27-MAR-2002; 2002WO-IB002163.
 XX
 XX 27-MAR-2001; 2001GB-00007658.
 XX
 XX (CHIR-) CHIRON SPA.
 XX (GENO-) INST GENOMIC RES.
 XX
 XX Masignani V, Tettelin H, Fraser C;
 XX
 XX WPI; 2003-040579/03.
 DR N-PSDB; ABX07751.
 DR
 DR
 DR
 DR
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX
 PS Claim 1; SEQ ID NO 4078; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AS556454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where
 CC the first primer is substantially complementary to the target sequence of
 CC and the second primer is substantially complementary to the target sequence
 CC of the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus

CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2469 proteins
 CC expressed by the identified coding regions from the genomic sequence.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX

SQ Sequence 210 AA;

Query Match 37.3%; Score 380.5; DB 6; Length 210;
 Best Local Similarity 41.0%; Pred. No. 3.le-33;
 Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;
 QY 1 MIEINDLKSPGVRIILWQGLSHKFLPGMTALTGASGSGKSTLLNCLGTLPKSPSGQLLV 60
 Db 1 MIELKQVSKSPGERELFSNLTPEAGKVYALIGSSGKTLTLMNMGKLE-PYDGTIFY 59
 QY 61 EDVDLLKLSRQRLYRKNTVGYLFQDYVALIPDRTVKFNQLAVEKHKWPEI-----PQ 114
 Db 60 RGKD---LANYKSSDFFRHELGYLEFQNGLIENSIENLKLIGLQKLSRSEORLRKQ 116
 QY 115 VLHVAVLESFE-EKVPFELSGEQORTALARVLLKNPRIILADPTGALDITNSLVEIA 173
 Db 117 ALEQVGLVYLDLDRIFELSGESQORVALAKIILKNPPFIILADEPTASIDPATSQLIMEI 176
 QY 174 LEADKCATVVVATHSPFPRESADTIKL 203
 Db 177 LLSURDDNRLIIITHNPAINWEMADEVPTM 206

RESULT 8
 ABP81469
 ID ABP81469 standard; protein; 210 AA.
 XX
 AC ABP81469;
 XX
 DT 04-MAR-2003 (first entry)
 XX
 XX Streptococcus pneumoniae polypeptide SEQ ID NO 386.
 DE
 DE Streptococcus pneumoniae; infection; otitis media; antibacterial;
 KW diagnosis; gene therapy.
 KW
 XX Streptococcus pneumoniae.
 XX
 XX WO200283855-A2.
 XX
 XX 24-OCT-2002.
 XX
 XX 12-APR-2002; 2002WO-US011524.
 XX
 XX 16-APR-2001; 2001US-0283948P.
 PR 18-APR-2001; 2001US-0284443P.
 XX
 XX (AMCY) AMERICAN CYANAMID CO.
 PA
 XX Zagursky RJ, Masi AW, Green BA, Chakravarti DN, Russell DP;
 PI Wothers JL;
 PI
 XX WPI; 2003-093010/08.
 DR N-PSDB; ABZ42317.
 DR
 XX New Streptococcus pneumoniae polynucleotides, useful for treating or
 PT preventing S. pneumoniae infections, or non-systemic diseases, e.g.
 PT otitis media, which are induced or exacerbated by S. pneumoniae.
 XX
 PS Claim 42; Page 637-638; 1091pp; English.

XX The invention relates to isolated polynucleotides (ABZ72147-ABZ42522) of
 CC a Streptococcus pneumoniae genomic sequence, a fragment or degenerate

variant of the polynucleotide or a nucleic acid sequence 95% identical to one of the polynucleotides. The S. pneumoniae polynucleotides and encoded polypeptides (ABP81299-ABP81674) are useful for treating or preventing S. pneumoniae infections or non-systemic diseases, e.g. otitis media, which are induced or exacerbated by S. pneumoniae. These are also useful for detecting S. pneumoniae in a biological sample or diagnosing S. pneumoniae infection in a subject. The polynucleotides have antibacterial activity and are useful in gene therapy

Query Match 37.3%; Score 380.5; DB 6; Length 210;
Best Local Similarity 41.0%; Pred. No. 3.1e-33;
Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;
QY 1 MIEINDLKSGFVRIWQGLSHKFLPGTMTALTGASGSGKSTLNLCLGTLDKPSGGILV 60
DB 1 MIEIKQVSKFGEELFNSLMTPEAGKVVALIGSSGSGKTLNMIGKLE-PYDTIFY 59
QY 61 EDVLLKSLTRKQRLYKNTVGYLFQDYALIPDRTVKFNQLQAVEKHWPDI-----PQ 114
DB 60 RGKD---LANYKSDFFRHELGFLYFQNGFLIENQSIENLKLGLIGCKLSRSRQRLRQK 116
QY 115 VLHVGLESPE-EKPVFELSGEQRTALARVLLKNPRIILADEPTGALDLTNSLVIEA 173
DB 117 ALEQVGLYLDKRIEFLSGSGEQRTALAKIILKNPFIILADEPTASIDPATISQLIMEI 176
QY 174 LRALADKATVVVATHSPFRESADTIKL 203
DB 177 LLSLRDNDRLIIATHNPATWEMADEVFTM 206

RESULT 9
ABB53309
ID ABB53309 standard; protein; 211 AA.
XX ABB53309;
XX 29-AUG-2003 (revised)
DT 16-MAY-2002 (first entry)
XX Lactococcus lactis protein yabE.
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX Lactococcus lactis; IL1403.
XX FR2807446-A1.
XX 12-OCT-2001.
XX 11-APR-2000; 2000FR-00004630.
PR 11-APR-2000; 2000FR-00004630.
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
PA Bolotine A, Sorokine A, Renault P, Ehrlich SD;
PI WPI; 2002-043418/06.
XX New nucleotide sequence useful in the identification of Lactococcus lactis and related species.
PT Claim 6; SEQ ID NO 11; 2504pp; French.

The present invention is related to a Lactococcus lactis nucleotide sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic acid sequence is useful in the detection and/or amplification of nucleic acid sequence, particularly to identify Lactococcus lactis or related species. The proteins of the invention are useful for the biosynthesis or biodegradation of a composition of interest. The invention helps research in lactic bacteria, particularly useful in the production of yogurt and

cheese. Note: The sequence data for this patent is based on equivalent patent WO200177334 (published 18-OCT-2001) which is available in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to standardise OS field)

Query Match 36.6%; Score 373; DB 5; Length 211;
Best Local Similarity 39.2%; Pred. No. 2.1e-32;
Matches 82; Conservative 42; Mismatches 79; Indels 6; Gaps 1;
QY 1 MIEINDLKSGFVRIWQGLSHKFLPGTMTALTGASGSGKSTLNLCLGTLDKPSGGILV 60
DB 1 MIEIEELTKSKYGHIFDKLNRIPGKQTAIYGTSGAGKSTLNLIGLIDYDGGKYF 60
QY 61 EDVLLKSLTRKQRLYKNTVGYLFQDYALIPDRTVKFNQLQAV-----EKHKWPEIPQ 114
DB 61 NGQFAPPFNSSIALKQRENKISYLFQNFALLEDETIEKNLSIALIYGRISKKEKKMKK 120
QY 115 VLHVGLESPEKPVFELSGEQRTALARVLLKNPRIILADEPTGALDLTNSLVIEAL 174
DB 121 LLLQVGHNRHLNTKYVLSGGEKQRTAARALLKESQLILADEPTGSLDTENRNEVIAL 180
QY 175 RALADKATVVVATHSPFRESADTIKL 203
DB 181 RQEVDKGKAVIVTHDSYLKVESDLVIEI 209

RESULT 10
ABU36156
ID ABU36156 standard; protein; 465 AA.
XX ABU36156;
XX 19-JUN-2003 (first entry)
DT Protein encoded by Prokaryotic essential gene #21683.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Mycoplasma pneumoniae.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA40026.
XX New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 64080; 1766pp; English.
PS The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC on a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 465 AA;

Query Match 35.6%; Score 362.5; DB 6; Length 465;
 Best Local Similarity 36.7%; Pred. No. 9.9e-31;
 Matches 77; Conservative 54; Mismatches 70; Indels 9; Gaps 4;
 QY 1 MIBINDLKK--SGV--RILWQLSHKFLPGTWTALTGASGSKSTLLNCLGTLTKPSSG 56
 Db 231 IIEELKNVYKTYINGVTNVLKGLDKLKAHDPIVILGSPGSKTLLNIISGMDRPPSSG 290
 QY 57 QILVEDVDLLKSTRKQRYKNTVGYLFQDYALIPDRTVKFNQLAVEKHKEWPE---I 112
 Db 291 SVVNVQEMICMDRQLTFRNRYVGYIFQYGLLENLVRENVVGANLQNPDKRINI 350
 QY 113 PQVLHVGLESFEEKPVFELSGGQQORTALARVLLKNPRIILADEPTGALDNLNSELVIE 172
 Db 351 DELEAVGMKHLQKLPNELSGGQQQVSTARAFAXNPILLIFGDEPTGALDLEMTQIVLK 410
 QY 173 ALPALADK-GATVVVATHSPFPRESADTII 201
 Db 411 QFLAKQRYKTYTIVITHNNLIAQLADLVI 440

RESULT 11

ID AAU45701

XX AAU45701 standard; protein; 258 AA.

XX AC AAU45701;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #6597.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.
 PR 02-JUN-2000; 2000US-0208841P.
 BR 07-JUL-2000; 2000US-0216747P.
 XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 PI WPI: 2001-616774/71.
 DR N-PSDB; AAS59526.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.

PS Claim 3; SEQ ID NO 6896; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 258 AA;

Query Match 35.2%; Score 359; DB 4; Length 258;
 Best Local Similarity 40.5%; Pred. No. 9.9e-31;
 Matches 85; Conservative 36; Mismatches 65; Indels 24; Gaps 3;

QY 12 GVRILWQLSHKP-----LPGTWTALTGASGSKSTLLNCLGTLTKPQS 54
 Db 34 GGSIVAGVSHNFSTTHRNRLVDDVTVMQPGVMTLLHGVSGSKTLLNINSGLLKPT 93
 QY 55 SGQILVEDVDLLKSTRKQRYKNTVGYLFQDYALIPDRTVKFNQLAVEKHKEWPEIQ 114
 Db 94 SGTVFEGSDIYSLSTSERODIRLNRMIFQHSLLIVDFTVRENVELLRVFGSGRSR 153
 QY 115 V-----LHVGLESFEEKPVFELSGGQQORTALARVLLKNPRIILADEPTGALDNLNSEL 169
 Db 154 VMAIEALERVGIAHLQDYPRLQSLSGGEAQRVGIARAIAGDRPILLADEPTGQLDRRSQM 213
 QY 170 VIEALRALA--DKGATVVVATHSPFPRESA 197
 Db 214 VFELLRALAEDEKGRTVLSSHSDTAKEYA 243

RESULT 12

ID ABM42220

XX ABM42220 standard; protein; 258 AA.

XX AC ABM42220;

XX 20-OCT-2003 (first entry)

XX Propionibacterium acnes transporter-related polypeptide #6896.

KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine.
 OS Propionibacterium acnes.
 XX WO2003033515-A1.
 XX 24-APR-2003.
 XX 11-OCT-2002; 2002WO-US032727.
 XX 15-OCT-2001; 2001US-00978825.
 XX (CORI-) CORIXA CORP.
 XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 FI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 FI Barth B, Vallie-Douglas J;
 XX WPI: 2003-381789/36.
 DR N-PSDB; ACF64455.
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
 FT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 FT or for stimulating an immune response specific for a P. acnes protein.
 XX Claim 3k; SEQ ID NO 6896; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridisation. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a specifically claimed polypeptide which is predicted
 CC to be encoded by an ORF (open reading frame) contained within the P.
 CC acnes polynucleotides of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcr_sequences

XX Sequence 258 AA;

Query Match 35.2%; Score 359; DB 6; Length 258;
 Best Local Similarity 40.5%; Pred. No. 9.9e-31;
 Matches 85; Conservative 36; Mismatches 65; Indels 24; Gaps 3;

QY 12 GVRILWGLSHKF-----LPGTWLTATGASGSKTLNCLGTLDKPS 54
 DB 34 GGSIVASGVSHNFSTHNRNLVLDVTVSMQPGVNTLLHGVSGSKTLLINMSGLLKPT 93
 QY 55 SQQLIVEDVLLKLSRQRLRYKNTVGYLFQDYALIPDRFTVKFNQLAVEKHKEWPEIQ 114
 DB 94 SGTVFEGSDIVSLSTSRDRLNRIGMIFOEHSLLIVDFTVRENVELILRVGFGRSR 153
 QY 115 V-----LHVGLEFEEKPFVSLGGECQRTALARVLLKPRILLADEPTGADLTNSEL 169

Db 154 VMATEALERVGIAHLQDRYPRQLSGGEARVGIARAIAGDRPILLADEPTGQLDRRNSQM 213
 QY 170 VIEALRALA--DKGATVVVATHSPLFRESA 197
 DB 214 VFELLRALAEDEKRGRTVVLSHSDPTAKEYA 243

RESULT 13

ADC97592
 ID ADC97592 standard; protein; 270 AA.

XX AC ADC97592;

XX DT 01-JAN-2004 (first entry)

XX DE E. faecium protein sequence SEQ ID 7219.

XX KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
 KW abdominal-pelvic infection.

XX OS Enterococcus faecium.

XX PN US6583275-B1.

XX PD 24-JUN-2003.

XX PF 30-JUN-1998; 98US-00107532.

XX PR 02-JUL-1997; 97US-0051571P.

XX PA 14-MAY-1998; 98US-0085598P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX PI Doucette-Stamm LA, Bush D;

XX DR WPI: 2003-799836/75.

XX DR N-PSDB; ADC93938.

XX New isolated nucleic acid derived from Enterococcus faecium encoding an
 FT Enterococcus faecium polypeptide useful for detection, prevention and
 FT treatment of a pathological condition resulting from a bacterial
 PT infection.

XX Example 1; SEQ ID NO 7219; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridising to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to a
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acids is useful for diagnosing pathological conditions
 CC resulting from E. faecium bacterial infection (e.g. urinary tract
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of Candida albicans -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating Enterococcus faecium infections. The present sequence represents
 CC one if the disclosed E. faecium proteins.

XX Sequence 270 AA;

Query Match 34.6%; Score 353; DB 7; Length 270;

Best Local Similarity 40.9%; Pred. No. 4.9e-30;

Matches 76; Conservative 36; Mismatches 72; Indels 2; Gaps 2;

QY 20 LSHKFLPTMTALTGASGSKSTLNCILGTLKPSQQLIVEDVLLKLSRQRLRYK 79

DB 63 ISFSIEKELVILGFGSGAKSTILNIGLGMPTDEGQIIDDIDTIAQFSDKQLTAYERT 122

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:34:05 ; Search time 19 Seconds
(without alignments)
551.583 Million cell updates/sec

Title: US-09-868-338-9
Perfect score: 1019
Sequence: 1 MIBINDLKXSFGRILWQGL.....VVVATHSPLFRESADTIKL 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	353	34.6	270	4	US-09-107-532A-7219
2	332	32.6	256	4	US-09-134-001C-4600
3	330.5	32.4	1049	4	US-09-252-991A-25336
4	329	32.3	672	4	US-09-489-039A-12101
5	324	31.8	257	4	US-09-134-000C-3925
6	322.5	31.6	286	4	US-09-627-376-15
7	321	31.5	653	4	US-09-543-681A-5411
8	320.5	31.5	239	4	US-09-134-000C-6167
9	315	30.9	226	4	US-09-134-000C-5979
10	308	30.2	286	4	US-09-540-236-2931
11	306.5	30.1	276	4	US-09-489-039A-13021
12	306	30.0	242	4	US-09-134-001C-3932
13	305.5	30.0	221	4	US-09-107-532A-6667
14	305.5	30.0	221	4	US-09-134-000C-6517
15	303	29.7	396	4	US-09-489-039A-10491
16	301.5	29.6	236	4	US-09-134-001C-4595
17	300.5	29.5	287	4	US-09-489-039A-7596
18	299.5	29.4	270	4	US-09-134-001C-5561
19	299	29.3	272	4	US-09-489-039A-9558
20	298	29.2	794	4	US-09-134-000C-5518
21	295	28.9	254	4	US-09-134-001C-4440
22	295	28.9	674	4	US-09-328-352-8094
23	294.5	28.9	262	4	US-09-134-000C-5469
24	294.5	28.9	274	4	US-09-252-991A-23361
25	294	28.9	255	4	US-09-107-532A-5335
26	292	28.7	244	4	US-08-519-573-2
27	292	28.7	244	4	US-08-519-573-4

ALIGNMENTS

RESULT 1

US-09-107-532A-7219
; Sequence 7219, Application US/09107532A
; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 7219:

SEQUENCE CHARACTERISTICS:

LENGTH: 270 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...270

SEQUENCE DESCRIPTION: SEQ ID NO: 7219:

US-09-107-532A-7219

Sequence 6978, Ap
Sequence 6466, Ap
Sequence 10373, A
Sequence 64, Appl
Sequence 64, Appl
Sequence 20657, A
Sequence 6078, Ap
Sequence 3731, Ap
Sequence 7794, Ap
Sequence 5226, Ap
Sequence 18, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 4327, Ap
Sequence 4444, Ap
Sequence 14256, A
Sequence 72, Appl
Sequence 74, Appl

```

Query Match      34.6%; Score 353; DB 4; Length 270;
Best Local Similarity 40.9%; Pred. No. 2.6e-32;
Matches 76; Conservative 36; Mismatches 72; Indels 2; Gaps 2;

QY 20 LSHKFLPGTWTALTGAGSGKSLNLCIGLMDKPSGQILVEDVLLKLSKQRLYRKN 79
DB 63 ISFSIERGELVILILPGSAGKSTLNLGMDPDDEGQIIIDTDIAQFSKDLTAIRYT 122
QY 80 TVGLFDYALIPDRTVKFNQLAVEKHWKEIP-QVILHAVGLESPPEKPFVFLSGEQ 138
DB 123 DVGFFQFYNLVNLTAKENVELATEVSPDLPDVEVLRQVLAHLNPPSLSGEGEQ 182
QY 139 RTALARVLLKNPRIILADEPTGALDLNLSSELVEALR-ALADKATVVVATHSPLFRESA 197
DB 183 RVSIARALAKNPKLLCDEPTGALDFETGKQVLLQNASRQHGNTVLIITHNSALAPIA 242
QY 198 DTIIKL 203
DB 243 DRVIHI 248

RESULT 2
US-09-134-001C-4600
; Sequence 4600, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4600
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4600

Query Match      32.6%; Score 332; DB 4; Length 256;
Best Local Similarity 37.6%; Pred. No. 6.3e-30;
Matches 80; Conservative 45; Mismatches 78; Indels 10; Gaps 5;

QY 1 MIEINDLKSPGVRILMQGL---SHKFLPGTWTALTGAGSGKSTLNLCLGTLDPKSSGQ 57
DB 7 ILTANQLSKVYGHKQKRALNNISFVSVEKGFVAVMGPSGSGKTTLLNLSIDTISGGT 66
QY 58 ILVEDVDLLKLSKQRLYRKNVGYLFDYALIPDRTVKFN--LQAVEK-HKW---PE 111
DB 67 VEVEGKEINKLSHKEVANFRQHLGFIPQDSVLPITLVKENIMPLSVQKFKYEMEQN 126
QY 112 IPQVLHAVGLESPPEKPFVFLSGEQQRTALARVLLKNPRIILADEPTGALDLNSELVI 171
DB 127 YKEVAEALGIYNLGNKYPSEISGGQQRTAAARAFVHKPTIIPADEPTGALDSKAQDLL 186
QY 172 EALRALADK-GATVVVATHSPLFRESADTIKL 203
DB 187 HRLEDNMQFNSTIMMVTDPSPSAAYABRVIML 219

RESULT 3
US-09-252-991A-25336
; Sequence 2536, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25336
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25336

Query Match      32.4%; Score 330.5; DB 4; Length 1049;
Best Local Similarity 37.3%; Pred. No. 7.9e-29;
Matches 81; Conservative 38; Mismatches 81; Indels 17; Gaps 4;

QY 1 MIEINDLKSPG-----VRILMQGLSHKFLPGTWTALTGAGSGKSTLNLCLGTLDPKPS 54
DB 396 LIELRDIRKRYGGNGTPEVEVL-KGVSLSIHAGEFVAIVGASGSGKSTLNMILGCLDRPS 454
QY 55 SQQILVEDVDLLKLSKQRLYRKNVGYLFDYALIPDRTVKFNQLQ-----AVEK 106
DB 455 SGSYHFAGHDVAELDSDEQAWLREAFVFGVYHLLIPSAQENVEPAIYAGIPASER 514
QY 107 HKWPEIPOVLHAVGLESPPEKPFVFLSGEQQRTALARVLLKNPRIILADEPTGALDLTN 166
DB 515 H--TPARALLRELGLAERTANRPHLSGGQQQVRSIARALMNGGHIIILADEPTGALDSHS 572
QY 167 SELVIEALRALADKATVVVATHSPLFRESADTIKL 203
DB 573 GAENVALLDELASQGHVVILITHDRDVAARAKRIIEV 609

RESULT 4
US-09-489-039A-12101
; Sequence 12101, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12101
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12101

Query Match      32.3%; Score 329; DB 4; Length 672;
Best Local Similarity 35.0%; Pred. No. 6e-29;
Matches 75; Conservative 44; Mismatches 83; Indels 12; Gaps 3;

QY 1 MIEINDLKSPF-----GVRILMQGLSHKFLPGTWTALTGAGSGKSTLNLCLGTLDPKPS 55
DB 30 LLELRDIRSYSPSGDGSVEVL-KGITLSIHAGEMVAIVGASGSGKSTLMTLGLCLDKETS 88
QY 56 QQILVEDVDLLKLSKQRLYRKNVGYLFDYALIPDRTVKFNQLAV-----EKHW 109
DB 89 GTYRVAGTDIAQDGDALARLREHFQIFQYHLSHLTAQNVVEPAVAGNERRARL 148
QY 110 PEIPQVLHAVGLESPPEKPFVFLSGEQQRTALARVLLKNPRIILADEPTGALDLTNSEL 169
DB 149 ARAHELLVELGLGERADYQPSQLSGGQQQVRSIARALMNGGVEVILADEPTGALDSHSGEE 208
QY 170 VIEALBALADKATVVVATHSPLFRESADTIKL 203

```

Db 209 VMAILHQLKAQGHVTIIVTHDPQVAAQAERIVEI 242

RESULT 5

US-09-134-000C-3925

; Sequence 3925, Application US/09134000C

; Patent No. 6617156

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-032

; CURRENT APPLICATION NUMBER: US/09/134,000C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/055,778

; PRIOR FILING DATE: 1997-08-15

; NUMBER OF SEQ ID NOS: 6812

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 3925

; LENGTH: 257

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-134-000C-3925

Query Match 31.8%; Score 324; DB 4; Length 257;

Best Local Similarity 35.1%; Pred. No. 5.3e-29;

Matches 72; Conservative 52; Mismatches 65; Indels 16; Gaps 4;

QY 1 MIEINDLKKSFGVRLWQ---GLSHKFLPGTWTALTGASGSKSTLLNCLGTLDPSSGQ 57

Db 9 MLEVQNLKKVYGNIEIKYEAALGKINLTVDQGEFTGIMGPGSGKSTLLNLLATIDPTDGE 68

QY 58 ILVEDVDLLKLSRKQRLYKNTVGYLFQDYALIPDRTVKFNL-----QLAVEKHK 108

Db 69 ILLNGKPNNLNQEQAKFRRTGELGFVFSFNLMPLTVEENIILPLTLDGKSVMKRQ 128

QY 109 WEIPQVLHAGVLESFEKVPFELSGEQOQTALARVLLKNPRIILADEPTGALDLTNS 168

Db 129 LAELSERL---GINHLKKRIABISGQQRVAVARAMTHHPQLLADDEPTGNTKSK 185

QY 169 LVIEALRAL-ADKATVAVATHSPL 192

Db 186 DVNGLLQQLNEEAATILMVTHDPL 210

RESULT 6

US-09-627-376-15

; Sequence 15, Application US/09627376

; Patent No. 6342385

; GENERAL INFORMATION:

; APPLICANT: Qi, Fengxia Caulfield, Page Chen, Ping

; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS

; FILE REFERENCE: UAB-17402/22

; CURRENT APPLICATION NUMBER: US/09/627,376

; CURRENT FILING DATE: 2001-05-30

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 15

; LENGTH: 246

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-627-376-15

Query Match 31.6%; Score 322.5; DB 4; Length 246;

Best Local Similarity 35.8%; Pred. No. 7.4e-29;

Matches 78; Conservative 41; Mismatches 74; Indels 25; Gaps 4;

QY 2 IEINDLKKSFGVRLWQLSHKL-----PGTWTALTGASGSKSTLLNCLGTLDPK 53

Db 4 IGVSHLKKVYKQ---EGTNEALXKDTTFSVQGEFTGIMGPGSGKSTLLNLLACMDYP 60

QY 54 SSGQILVEDVDLLKLSRKQRLYKNTVGYLFQDYALIPDRTVKFNLQAV----- 104

Db 61 SSGHIIFNNTVQLEKVKDEBAAVFRSRHIGFIFQNFULLNFIENKDNLLIPVILSGSKVNS 120

QY 105 -EKHKWPEIFQVLHAGVLESFEKVPFELSGEQOQTALARVLLKNPRIILADEPTGALD 163

Db 121 YEK---RLRDLAAVVGIESLLSKYPYELSGGQOQLAIARALIMNPDLLADEPTQOLD 176

QY 164 LTRSELVIEALRALADKATVAVATHSPLPRESADTII 201

Db 177 SKTSORIILNLSINAKRKILMVTHSPKAAVSANRVL 214

RESULT 7

US-09-543-681A-5411

; Sequence 5411, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 5411

; LENGTH: 653

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-5411

Query Match 31.5%; Score 321; DB 4; Length 653;

Best Local Similarity 35.8%; Pred. No. 4.8e-28;

Matches 77; Conservative 46; Mismatches 78; Indels 14; Gaps 5;

QY 1 MIEINDLKKSFP---GVRILWQLSHKLPGTWTALTGASGSKSTLLNCLGTLDPSSG 56

Db 10 LLENNVSRLYNGEEDTVLVDISLTINAGEMVAITGASGSKSTLLNCLGTLDPSSG 69

QY 57 QILVEDVDLLKLSRKQRLYKNTVGYLFQDYALIPDRTVKFNLQ-----AVEKHK 109

Db 70 EYKAGOCVADMESDQLAALRREHFGIFQRYHLMAHLTAEQNVEIPAIYAGKSTEQRK- 128

QY 110 PEIPQVLHAGVLESFEKVPFELSGEQOQTALARVLLKNPRIILADEPTGALDLTNS 168

Db 129 ERARALLTRLGLAERIHYP-SQLSGGQOQRVSTARALMNGGEVILADEPTGALDSQSK 187

QY 169 LVIEALRALADKATVAVATHSPLPRESADTIIKL 203

Db 188 EVMAILQQLNQGHVTIIVTHDPLIAQADRILIEI 222

RESULT 8

US-09-134-000C-6167

; Sequence 6167, Application US/09134000C

; Patent No. 6617156

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-032

; CURRENT APPLICATION NUMBER: US/09/134,000C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/055,778

; PRIOR FILING DATE: 1997-08-15

; NUMBER OF SEQ ID NOS: 6812

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 6167

; LENGTH: 239

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

US-09-134-000C-6167

```
Query Match 31.5%; Score 320.5; DB 4; Length 239;
; Number of Seq ID NOS: 3840
; SEQ ID NO 2931
; Length: 286
; Type: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-2931

Query Match 30.2%; Score 308; DB 4; Length 286;
Best Local Similarity 34.2%; Pred. No. 4.4e-27;
Matches 75; Conservative 50; Mismatches 76; Indels 18; Gaps 5;

QY 2 IEINDLKSGFVRILWQGLSHKFLPGTWTALTGASGSGKSTLLNCLGTLDPKSSGQILVE 61
Db 38 LDLQDIHKSYSLEVLKGVSLTAYDGDVISITLSSGSGKSTLLNCLGTLDPKSSGQILVE 97
QY 62 DVDLLKSTRKQRLYR-----KNTVGYLFQDYALIPDRTVKFNL-----OLAVEK 106
Db 98 Q-ETILKSKHGELVAADNKOLEKLRSGKIGVFQFNLPKHTLQNIIEGPTCVLGISK 156
QY 107 -HKWPEIQVLHVGLESFEEKPVFELSGGQOQTALARVLLKNPRIILADEPTGALDLT 165
Db 157 ADAIKDAERLLTKVGLLDKDAYDPNLSGGQQRVAIARSLAMQPQVLLDEPTGALDPE 216
QY 166 NSELVIEALRALADKATVVVATHSPFLF-RESADTIKL 203
Db 217 LVNEVLAVMRELADEGRITMLIVTHEMFAREVSKVVEL 255

RESULT 11
US-09-489-039A-13021
; Sequence 13021, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13021
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13021

Query Match 30.1%; Score 306.5; DB 4; Length 276;
Best Local Similarity 36.5%; Pred. No. 6.2e-27;
Matches 76; Conservative 40; Mismatches 83; Indels 9; Gaps 4;

QY 2 IEINDLKSGFVRILWQGLSHKFLPGTWTALTGASGSGKSTLLNCLGTLDPKSSGQILVE 61
Db 28 ITRDVAKRYGDHQLVNAINLQVESGEVAILGPGSGKSTLLNCLGTLDPKSSGQILVE 87
QY 62 DVDLLKSTRKQRLYRKNVTGYLFQDYALIPDRTVKFNLOLAVEK-HKWPEIP-----Q 114
Db 88 GKPTSRLSGSGSALRQR-SRVGFVQQFNLYAHLTAQENITLALERVHVGWGSAAARALA 146
QY 115 VLHVGLESFEEKPVFELSGGQOQTALARVLLKNPRIILADEPTGALDLTNSSELVIEAL 174
Db 147 LLRQVGLLEKAQOMPAQLSGGQQRVAIARALASSPOIILFDEPTGALDPEMIGVLM 206
QY 175 RALADKATVVVATHSPFLF-RESADTI 201
Db 207 KTLAHSGITMLVVTHEMQFAREIADRVV 234

RESULT 12
US-09-134-001C-3832
; Sequence 3832, Application US/09134001C
```

```
Query Match 31.5%; Score 320.5; DB 4; Length 239;
; Number of Seq ID NOS: 3840
; SEQ ID NO 2931
; Length: 286
; Type: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-2931

Query Match 30.2%; Score 308; DB 4; Length 286;
Best Local Similarity 34.2%; Pred. No. 4.4e-27;
Matches 75; Conservative 50; Mismatches 76; Indels 18; Gaps 5;

QY 2 IEINDLKSGFV-----RIWQGLSHKFLPGTWTALTGASGSGKSTLLNCLGTLDPKSS 55
Db 14 LIELSDINKFYPVGEKHLVL-KELNLTIIHQGEFILLGSGSGKSTLLNCLGTLDPKSS 72
QY 56 GQILVEDVLLKSTRKQRLYRKNVTGYLFQDYALIPDRTVKFNLOI-----AVEK 106
Db 73 GSYHFGTDVSKLSENKSAFRNLYGLFIQQFFLNSLVNSQVQLPCVYEGKGRAEK 132
QY 107 HKWPEIQVLHVGLESFEEKPVFELSGGQOQTALARVLLKNPRIILADEPTGALDLTN 166
Db 133 KAIAE--KYLKIVGLETKAKSKVTELSSGGQQRVAIARSLVNDPFLIMADEPTGALDSET 190
QY 167 SELVIEALRALADKATVVVATHSPFLRESADTIKL 203
Db 191 GTEIMELLKELNEQKTIIVMTHTDEDMKKVASRVIHM 227

RESULT 9
US-09-134-000C-5879
; Sequence 5879, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134.000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5879
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5879

Query Match 30.9%; Score 315; DB 4; Length 226;
Best Local Similarity 38.3%; Pred. No. 4.8e-28;
Matches 70; Conservative 36; Mismatches 71; Indels 6; Gaps 1;

QY 27 GTMTALTGASGSGKSTLLNCLGTLDPKSSGQILVEDVLLKSTRKQRLYRKNVTGYLFQ 86
Db 33 GEFLAINGPSGSGKSTLLNCLGTLDPKSSGQILVEDVLLKSTRKQRLYRKNVTGYLFQ 92
QY 87 DYALIPDRTVKFNLOLAV-----EKHWPEIQVLHVGLESFEEKPVFELSGGQOQT 140
Db 93 NFSLIETLVEENIELPLLYSGLTPEAKDRVHEVLTQVGLPKGKHPKQLSGGQOQRV 152
QY 141 ALARVLLKNPRIILADEPTGALDLTNSSELVIEALRALADKATVVVATHSPFLRESADTI 200
Db 153 AIAARVNRFSFIIDAEPTGALDLSKTSIEITLTFQQLNNEGVTIILVTHDEETIEYCNRL 212
QY 201 IKL 203
Db 213 IKV 215

RESULT 10
US-09-540-236-2931
; Sequence 2931, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540.236
```


Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3832
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3832

Query Match 30.0%; Score 306; DB 4; Length 242;
Best Local Similarity 33.5%; Pred. No. 5.8e-27;
Matches 73; Conservative 44; Mismatches 73; Indels 28; Gaps 4;

Qy 1 MIEINDLKSGFVRIWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
Db 3 VINIKNLKFGANEVLRDLNLTVEKEVVAIGPSGSGKSTLLRCWNLLDVPKGVIF 62

Qy 61 EDVDLLKSTRKQRLRYKNTVGYLFQDYALIPDRTVKFNQLAVEKHKWPPIQVILHVG 120
Db 63 EDNELTQHNHLDNLQK-MGMVFQNFNLFPHKVIENWMLA-----PLLLHKDS 111

Qy 121 LSGFEERPFV-----ELSGGEQORTALARVLLKNPRIILADEPTGALDL 164
Db 112 KDQLKEKALVLEKVGKLDKADSYFQLSGQQRVAIARALAMEPDVWLFDEPTSLDLP 171

Qy 165 TNSLVIEALRALADKATGVVATHSPLF-RESADTII 201
Db 172 EVVGDVLKVMRQLANESMTWIVTHENMFAKEISDKV 209

RESULT 13
US-09-107-532A-6667
; Sequence 6667, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6667:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5) LOCATION 1...221
; SEQUENCE DESCRIPTION: SEQ ID NO: 6667:
US-09-107-532A-6667

Query Match 30.0%; Score 305.5; DB 4; Length 221;
Best Local Similarity 37.5%; Pred. No. 5.8e-27;
Matches 78; Conservative 45; Mismatches 72; Indels 13; Gaps 6;

Qy 1 MIEINDLKSGF---VRIWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQ 57
Db 6 ILEIKNAVSYANSKEVL-SGVNQKPELGKPYAIVGSGTGKSTLLSLLAGLDKPGTGK 64

Qy 58 ILVEDVDLLKSTRKQRLRYKNTVGYLFQDYALIPDRTVKFNQLAVEKHKWPPIQVILH 117
Db 65 ILFKNED---IQNGYSNHRKNISLVFQNYNLIDYLSPIENIRLVNKSAD----ESILF 117

Qy 118 AVGLESPF-EKVPFELSGGEQORTALARVLLKNPRIILADEPTGALDITNSLVIEALRA 176
Db 118 ELGLDKKQIKENVKMLSGGQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIINILKT 177

Qy 177 LA-DKGATVVVATHSPLPRESADTIIKL 203
Db 178 LAKDENKCVIVVTHSKEVADSADIILEL 205

RESULT 14
US-09-134-000C-6517
; Sequence 6517, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6517
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6517

Query Match 30.0%; Score 305.5; DB 4; Length 221;
Best Local Similarity 37.5%; Pred. No. 5.8e-27;
Matches 78; Conservative 45; Mismatches 72; Indels 13; Gaps 6;

Qy 1 MIEINDLKSGF---VRIWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQ 57
Db 6 ILEIKNAVSYANSKEVL-SGVNQKPELGKPYAIVGSGTGKSTLLSLLAGLDKPGTGK 64

Qy 58 ILVEDVDLLKSTRKQRLRYKNTVGYLFQDYALIPDRTVKFNQLAVEKHKWPPIQVILH 117
Db 65 ILFKNED---IQNGYSNHRKNISLVFQNYNLIDYLSPIENIRLVNKSAD----ESILF 117

Qy 118 AVGLESPF-EKVPFELSGGEQORTALARVLLKNPRIILADEPTGALDITNSLVIEALRA 176

Db 118 ELGLDKQIRKRVNWKLSGGQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIINILKT 177

Qy 177 LA-DKATVVVATHSPILFRESADTIKL 203

Db 178 LAKDRKCVIVVTHSKEVADSADIILEL 205

RESULT 15

US-09-489-039A-10491
; Sequence 10491, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10491
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10491

Query Match 29.7%; Score 303; DB 4; Length 396;
Best Local Similarity 35.3%; Pred. No. 2.7e-26;
Matches 73; Conservative 45; Mismatches 73; Indels 16; Gaps 5;
Qy 1 MIEINDLKSGVRILMCGLSHKLPETMTALTGASGSGKSTLNCICLTGIDKPSGQILV 60
Db 38 LLEIRNLTKSFDGQHAVDDVSLTIYKGEIFALLGASGSGKSTLLRMLAGFEQPTAGQIML 97
Qy 61 EDVLLKLTQRKQRLYRKNVGYLFDYALIPDRTVKNQLAVEKHKEP-----IPQ 114
Db 98 DGVDLARYPP-----YQR-PINMMFQSYALFPFHMTVEQNIAFGKQDLPKABITARVQE 151
Qy 115 VLHVGLESFEKPVFELSGEQRTALARVLLKNPRIILADEPTGALD-LTNSSELVIEA 173
Db 152 MLALVHMGEFAKRPKHQSLSGQRQVALARSLAKRPKLLLDDEMGALDKKLRDRVQLEV 211
Qy 174 LRALADKATVVVATHSPILFRESADTI 200
Db 212 VDILERVGVTCVMVTHD---OEEAMTM 235

Search completed: July 27, 2004, 16:37:19
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:36:31 ; Search time 45 Seconds
(without alignments)
1412.673 Million cell updates/sec

Title: US-09-868-338-9

Perfect score: 1019
Sequence: 1 MEINDLKSGFVRILMQSL.....VVVATHSPFLRESADTIKL 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 31154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	399	39.2	213	16	US-10-474-776-389
2	380.5	37.3	210	16	US-10-474-776-386
3	371.5	36.5	248	14	US-10-156-761-11171
4	362.5	35.6	465	12	US-10-282-122A-64080
5	353	34.6	466	12	US-10-282-122A-63428
6	344.5	33.8	227	12	US-10-282-122A-67001
7	339.5	33.3	222	9	US-09-738-626-4456
8	339.5	33.3	222	12	US-10-627-476-442
9	339	33.3	648	9	US-09-815-242-13955
10	337	33.1	282	12	US-10-282-122A-44079
11	336.5	33.0	315	14	US-10-156-761-12964
12	336	33.0	236	9	US-09-815-243-13482
13	336	33.0	249	9	US-09-815-243-5642
14	336	33.0	252	9	US-09-815-242-12706
15	333.5	32.7	231	14	US-10-238-075-1256

16	332	32.6	252	12	US-10-282-122A-70852	Sequence 70852, A
17	330.5	32.4	663	9	US-09-815-242-11869	Sequence 11869, A
18	330.5	32.4	663	14	US-10-246-330-20	Sequence 20, Appl
19	330	32.4	241	12	US-10-282-122A-60354	Sequence 60354, A
20	329.5	32.3	248	12	US-10-282-122A-60354	Sequence 62354, A
21	329.5	32.3	248	12	US-10-282-122A-64477	Sequence 64477, A
22	329.5	32.3	261	14	US-10-156-761-14933	Sequence 14933, A
23	328	32.2	252	12	US-10-282-122A-71920	Sequence 71920, A
24	325.5	31.9	247	15	US-10-369-493-10091	Sequence 10091, A
25	324.5	31.8	233	9	US-09-815-242-13827	Sequence 13827, A
26	324	31.8	248	12	US-10-282-122A-53217	Sequence 53217, A
27	324	31.8	257	9	US-09-815-242-10577	Sequence 10577, A
28	323.5	31.7	255	12	US-10-282-122A-61025	Sequence 61025, A
29	322.5	31.6	246	13	US-10-047-676A-15	Sequence 15, Appl
30	321	31.5	271	9	US-09-815-242-13350	Sequence 13350, A
31	321	31.5	271	9	US-09-815-242-13690	Sequence 13690, A
32	320.5	31.5	256	12	US-10-282-122A-51558	Sequence 51558, A
33	315	30.9	242	12	US-10-282-122A-70384	Sequence 70384, A
34	314	30.8	240	12	US-10-282-122A-58935	Sequence 58935, A
35	314	30.8	379	12	US-10-282-122A-46319	Sequence 46319, A
36	312.5	30.7	243	12	US-10-282-122A-51603	Sequence 51603, A
37	312.5	30.7	262	14	US-10-156-761-11088	Sequence 11088, A
38	312	30.6	242	12	US-10-282-122A-70607	Sequence 70607, A
39	311	30.5	230	9	US-09-738-626-6814	Sequence 6814, Ap
40	311	30.5	230	12	US-10-627-476-334	Sequence 334, App
41	310.5	30.5	245	12	US-10-282-122A-52619	Sequence 52619, A
42	310	30.4	222	12	US-10-282-122A-54068	Sequence 54068, A
43	310	30.4	247	12	US-10-282-122A-69667	Sequence 69667, A
44	309.5	30.4	228	9	US-09-815-242-10135	Sequence 10135, A
45	309.5	30.4	236	9	US-09-738-626-6304	Sequence 6304, Ap

ALIGNMENTS

RESULT 1
US-10-474-776-389
; Sequence 389, Application US/10474776
; Publication No. US00040110181A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING POLYPEPTIDES
; FILE REFERENCE: AM100649-PCT
; CURRENT APPLICATION NUMBER: US/10/474,776
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 752
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 389
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-474-776-389

Query Match 39.2%; Score 399; DB 16; Length 213;
Best Local Similarity 42.9%; Pred. No. 3.3e-34;
Matches 90; Conservative 40; Mismatches 70; Indels 10; Gaps 3;

Qy	1	MEINDLKSGFVRILMQSLPCTMTALTGASGSGKSTLLNCITGLDKPSSGOILV	60
Db	1	MIDIQLEKFPNDRAIFSLGNLKLKRGKVALTGKSGKSTLLNLTGKLEKIDGGRLV	60
Qy	61	EDVLLKLSTKQRLYKNTVGVLFQDYALTPORTVFNQLQAVKHKWPEIPQVLHVG	120
Db	61	QKDLKTIPTRE---YFRDQGVLFQNFGLLENQSIKENLDLGFVGQKISKVERLERQVG	117
Qy	121	LESFE-----EKPVPFELSGGEOQTALARVLKNPRILABPTGALDITNSSELVIA	173
Db	118	ALEKVNGLDLEQKIYTLGSGEQRVALAKTIKNPLPLADEPTAALDPENSEEVMNL	177
Qy	174	LRLADKATVAVATHSPFLRESADTIKL	203
Db	178	LVDLKDNRILIIATHNPLVWNADEILDM	207

```
RESULT 2
US-10-474-776-386
; Sequence 386, Application US/10474776
; Publication No. US20040110181A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NOVEL STREPTOCOCCUS PNEUMONIAE OPEN READING FRAMES ENCODING POLYH
; FILE REFERENCE: AM100649-PCT
; CURRENT APPLICATION NUMBER: US/10/474,776
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 752
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 386
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-474-776-386

Query Match      37.3%; Score 380.5; DB 16; Length 210;
Best Local Similarity 41.0%; Pred. No. 3.1e-32;
Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;

QY 1 MIEINDLKSPGVRILWQGLSHKFLPCTMTALTGASGSGKSTLNCGLTLDKPSGQILV 60
DB 1 MELKQVSKSPGERELFNSLMTPEAGKVVALIGSGSGKSTLNMIGKLE-PYDGTIFY 59
QY 61 EDVLLKSLSTRKQRLYKNTVGYLFQDYALIPDRTVKFNQLQAVEKHKWEI-----PQ 114
DB 60 RGKD---LANYKSDFFRHEILGYLFQNFGLTIENOSIENKLGICQKLSRSQRURQKQ 116
QY 115 VLHVGLESFPEKVPFELSGEQQRTALARVLLKNPRIILADEPTGALDNLNSELVIEA 173
DB 117 ALEQVGLVYLDKRIEFLSGESQORVALAKIILKNPFFILADEPTASIDPATSQLIMEI 176
QY 174 LRALADKATVVVATHSPPLFRESADTIKL 203
DB 177 LLSLRDNNRLIIATHNPAINEMADEVFTM 206

RESULT 3
US-10-156-761-11171
; Sequence 11171, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11171
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11171

Query Match      36.5%; Score 371.5; DB 14; Length 248;
Best Local Similarity 41.0%; Pred. No. 3.6e-31;
Matches 87; Conservative 37; Mismatches 77; Indels 11; Gaps 4;

QY 1 MIEINDLKSPGVRILWQGLSHKFLPCTMTALTGASGSGKSTLNCGLTLDKPSGQILV 56
DB 9 IOLRSVSRRYGAGGAVTALDQ-VSLAFPRGTFTAVMGPSGSGKSTLLQCAAGLDRTSG 67
QY 57 QILYEDVDLLKSLSTRKQRLYKNTVGYLFQDYALIPDRTVKFN-----LQAVEKHKWEI 112
DB 68 SVTVGDTLTKLSETKTLFLRRDRIGFVFOAFNLLPSLTAEQNVALPRLAGRRPRTEV 127
QY 113 PVLHVGLESFPEKVPFELSGEQQRTALARVLLKNPRIILADEPTGALDNLNSELVIE 172
DB 128 REVLAQVGLGDRAGHRPTMSGGQORVALARALITRPDLVFGDEPTGALDSQTSREVL 187
QY 173 ALRALAD-KGATVVVATHSPPLFRESADTIKL 203
DB 188 LLRGVMDSEGQTVIMVTHDPVAASYADRVVFL 219

RESULT 4
US-10-282-122A-64080
; Sequence 64080, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64080
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Mycoplasma pneumoniae
US-10-282-122A-64080

Query Match      35.6%; Score 362.5; DB 12; Length 465;
Best Local Similarity 36.7%; Pred. No. 8.1e-30;
Matches 77; Conservative 54; Mismatches 70; Indels 9; Gaps 4;

QY 1 MIEINDLKSPGVRILWQGLSHKFLPCTMTALTGASGSGKSTLNCGLTLDKPSGQILV 56
DB 9 IOLRSVSRRYGAGGAVTALDQ-VSLAFPRGTFTAVMGPSGSGKSTLLQCAAGLDRTSG 67
QY 57 QILYEDVDLLKSLSTRKQRLYKNTVGYLFQDYALIPDRTVKFN-----LQAVEKHKWEI 112
DB 68 SVTVGDTLTKLSETKTLFLRRDRIGFVFOAFNLLPSLTAEQNVALPRLAGRRPRTEV 127
QY 113 PVLHVGLESFPEKVPFELSGEQQRTALARVLLKNPRIILADEPTGALDNLNSELVIE 172
DB 128 REVLAQVGLGDRAGHRPTMSGGQORVALARALITRPDLVFGDEPTGALDSQTSREVL 187
QY 173 ALRALAD-KGATVVVATHSPPLFRESADTIKL 203
DB 188 LLRGVMDSEGQTVIMVTHDPVAASYADRVVFL 219
```

Db 231 IIELKNVYKITNGVTNNAVLKGLDKLKAHDFIVILGPGSGKTTLLNIISGMDRPSG 290
 QY 57 QILVEDVLLKSTRKQRLYKNTVGVLFQDYALIPDRVTKENLQLAVERKWP- ---I 112
 Db 291 SVVNGQENICMDROLTRFRNYGYIFQGLPLNLTRENVGVANLQNPDKRNI 350
 QY 113 PQVLHAGVLESFEEKVPFELSGEQQTALARVLLKNPRIILADBTGALDITNSBELVIE 172
 Db 351 DELLEAVGMKHLQKLPNELSGQQQRVSIARAFAPKNPLLIFGDEPTGALDLEMTQIVLK 410
 QY 173 ALRALADK-GATVVVATHSPRESADTII 201
 Db 411 QFLAIRQRYKTTWIVTTHNLIQAQLADLVI 440

RESULT 5
 US-10-282-122A-63428
 ; Sequence 63428, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 63428
 ; LENGTH: 466
 ; TYPE: PRT
 ; ORGANISM: Mycoplasma genitalium
 US-10-282-122A-63428

Query Match 34.6%; Score 353; DB 12; Length 466;
 Best Local Similarity 39.2%; Pred. No. 8.5e-29;
 Matches 73; Conservative 42; Mismatches 61; Indels 10; Gaps 3;
 QY 21 SHKPLGTMALGASGSGKSTLLNCLGTLDPKSPGQILVEDVLLKSTRKQRLYKNT 80
 Db 261 SHDFI-----VILGPGSGGKTTLLNIISGMDRASSGVINGYNNICLNDKRLTKFRQY 315

QY 81 VGYLFQDYALIPDRVTKENLQLAVERKWP- ---EIPQVLHAGVLESFEEKVPFELSGE 136
 Db 316 VGYLFQDYALIPDRVTKENLQLAVERKWP- ---EIPQVLHAGVLESFEEKVPFELSGE 375
 QY 137 QORTALARVLLKNPRIILADBTGALDITNSBELVIEALRALADK-GATVVVATHSPLE 195
 Db 376 QQRVSIARAFAPKNPLLIFGDEPTGALDLEMTQIVLKQFLAKRYQTTMIIVTHNNLIAN 435
 QY 196 SADTII 201
 Db 436 LADLVI 441

RESULT 6
 US-10-282-122A-67001
 ; Sequence 67001, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 67001
 ; LENGTH: 227
 ; TYPE: PRT
 ; ORGANISM: Pasteurella multocida
 US-10-282-122A-67001

Query Match 33.8%; Score 344.5; DB 12; Length 227;
 Best Local Similarity 40.1%; Pred. No. 2.5e-28;
 Matches 83; Conservative 34; Mismatches 85; Indels 5; Gaps 3;
 QY 1 MIEINDLKSGFVILLWQGLSHKPLPGTMTALTGASGSGKSTLLNCLGTLDPKSPGQILV 60
 Db 1 MIETQHLKRFQGVQTALEDINIQIKEGFVAIMGASGSGKTTLMNLTGLDTASGKVL 60
 QY 61 EDVLLKLLSTRKQRLYKNTVGYLFQDYALIPDRVTKENLQLAVERKWP- ---IPOVLH 117

Db 61 DGVDAQDLBGRFRFAEKIGLVFQFHLIPYLTALENVMLACHYSHVIDEAAKAVLIQ 120
 Qy 118 AVGL-ESFEKVPFELSGGQORTALARVLKPNRIILADEPTGALDNLSELVIEA 176
 Db 121 QVGLAHFEDFRP-SQLSGGQQRVCIAIALVNPFPVIFADEPTGNLDEKNEQLVLDLVT 179
 Qy 177 LADKATVWVATHSPFPRESADTIKL 203
 Db 180 LNOQGRVWVWTHNPESKLADRTIFL 206

RESULT 7

US-09-738-626-4456
 ; Sequence 4456, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIALI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAKKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO 4456
 ; LENGTH: 222
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-4456

Query Match 33.3%; Score 339.5; DB 9; Length 222;
 Best Local Similarity 38.1%; Pred. No. 8.3e-28;
 Matches 80; Conservative 36; Mismatches 83; Indels 9; Gaps 3;
 Qy 2 IEINDLKKSPG-----VRILWQGLSHKFLPFTMTALTGASGSGKSTLLNCLGTLDPKPSG 56
 Db 8 LELQNISCAFGEGPRHVSAL-NNVSLAVNFGELVAMGPGSGKSTLLNVAGLQRTATSG 66
 Qy 57 QILVEDVDLLKSTRKQRLYRKNVTGYLFQDYALIPDRTVKFNQLAVE---KHKWPEIP 113
 Db 67 HVLIDGASADLNKRAAETRRRHIGVIFQNYNLVPTLTGVNGLPFLDGTDRQAVA 126
 Qy 114 QVLHVGLESFEEKVPFELSGGQORTALARVLKPNRIILADEPTGALDNLSELVIEA 173
 Db 127 IALAEVLEGFDPRFPEISGGQQRVAIARALIGPRKILLADEPTGALDNLSELVIEA 186
 Qy 174 LRALADKATVWVATHSPFPRESADTIKL 203
 Db 187 LRQRIDSGAAGLLVTHPRFAAWADRTIML 216

RESULT 8

US-10-627-476-442
 ; Sequence 442, Application US/10627476
 ; Publication No. US20040030116A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pompejus, Mark
 ; APPLICANT: Krogger, Burkhard
 ; APPLICANT: Schoder, Hartwig

; APPLICANT: Zelder, Oskar
 ; APPLICANT: Haberhauer, Gregor
 ; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
 ; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
 ; FILE REFERENCE: BGI-125CPCN
 ; CURRENT APPLICATION NUMBER: US/10/627,476
 ; CURRENT FILING DATE: 2003-07-25
 ; PRIOR APPLICATION NUMBER: 09/602,787
 ; PRIOR FILING DATE: 2000-06-23
 ; PRIOR APPLICATION NUMBER: USSN 60/141031
 ; PRIOR FILING DATE: 1999-06-25
 ; PRIOR APPLICATION NUMBER: DE 19931454.3
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: DE 19931478.0
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: DE 19931563.9
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: DE 19932122.1
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932124.8
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932125.6
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932128.0
 ; PRIOR FILING DATE: 1999-07-09
 ; PRIOR APPLICATION NUMBER: DE 19932180.9
 ; PRIOR FILING DATE: 1999-07-09
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 678
 ; SEQ ID NO 442
 ; LENGTH: 222
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-10-627-476-442

Query Match 33.3%; Score 339.5; DB 12; Length 222;
 Best Local Similarity 38.1%; Pred. No. 8.3e-28;
 Matches 80; Conservative 38; Mismatches 83; Indels 9; Gaps 3;
 Qy 2 IEINDLKKSPG-----VRILWQGLSHKFLPFTMTALTGASGSGKSTLLNCLGTLDPKPSG 56
 Db 8 LELQNISCAFGEGPRHVSAL-NNVSLAVNFGELVAMGPGSGKSTLLNVAGLQRTATSG 66
 Qy 57 QILVEDVDLLKSTRKQRLYRKNVTGYLFQDYALIPDRTVKFNQLAVE---KHKWPEIP 113
 Db 67 HVLIDGASADLNKRAAETRRRHIGVIFQNYNLVPTLTGVNGLPFLDGTDRQAVA 126
 Qy 114 QVLHVGLESFEEKVPFELSGGQORTALARVLKPNRIILADEPTGALDNLSELVIEA 173
 Db 127 IALAEVLEGFDPRFPEISGGQQRVAIARALIGPRKILLADEPTGALDNLSELVIEA 186
 Qy 174 LRALADKATVWVATHSPFPRESADTIKL 203
 Db 187 LRQRIDSGAAGLLVTHPRFAAWADRTIML 216

RESULT 9
 US-09-815-242-13995
 ; Sequence 1395, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A

```

RESULT 10
US-10-282-122A-44079
; Sequence 44079, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335

```

```

RESULT 11
US-10-156-761-12964
; Sequence 12964, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12964
; LENGTH: 315
; TYPE: PRT
; ORGANISM: streptomyces avermitilis
US-10-156-761-12964

```

Query Match	33.0%;	Score	336.5;	DB 14;	Length	315;			
Best Local Similarity	37.7%;	Pred. No.	2.8e-27;						
Matches	80;	Conservative	39;	Mismatches	82;	Indels	11;	Gaps	4;

QY 2 IEINDLKSPG-----VRILWQGLSHKPLPGTMTALTGASGSGKSTLLNCLGTLDPKSSG 56
DB 71 VELRGVRRYGRGSAVHAL-RGVDLALARGFTAVMSPSGSKSTFLQCAAGLDPSGG 129
QY 57 QILVEDVLLKSTRKQRYKNTVGYLFQDVALIPDRTVFN-----LQAVEKHKWPFI 112
DB 130 TVHLGTDITLSENKLTALRRSLRGVFAENLPLSLTVEQNVVLPRLFLAGHRPDRRA 189
QY 113 POVLHAGVLESPEEKVPFELSGGEQORTALARVLLKNPRIILADEPTGALDITNSSELVI 172
DB 190 AEVLQVGLQDKRRRPGQLSGGQORVAIAALVTRDVFVFADEPTGALDITTAAILG 249
QY 173 ALRALAD-KGATVVVATHSPRESADTIKL 203
DB 250 LLRQAVDVHGATVVMVTHDFTAAANADVLF 281

RESULT 12

US-09-815-242-13482
; Sequence 13482, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13482

; LENGTH: 236

; TYPE: PR1

; ORGANISM: Streptococcus pneumoniae

US-09-815-242-13482

Query Match 33.0%; Score 336; DB 9; Length 236;

Best Local Similarity 40.5%; Pred. No. 2,1e-27;

Matches 75; Conservative 38; Mismatches 58; Indels 14; Gaps 3;

QY 27 GTWTALTGASGSKSTLLNCLGTLDPKSSGQILVEDVLLKSTRKQRYKNTVGYLFQ 86

DB 36 GEIVITILGASGSKSTLLNCLGTLDPKSSGQILVEDVLLKSTRKQRYKNTVGYLFQ 95

QY 87 DYALIPDRTVKFNQLAVEKHKWPFI-----POVLHAGVLESPEEKVPFELSGGEQOR 139

DB 96 FYNLVNLTAKENVELA-----SEIVTALNSDQVLTVDVGLAHLNFPALSGGQOR 149

QY 140 TALARVLLKNPRIILADEPTGALDITNSSELVIALRALA-DKGATVVVATHSPRESAD 198

DB 150 VSTARAVAKNPKILLCDDEPTGALDYQTKGVKILQDMSROKGATVIVTHNGALAPIAD 209

QY 199 TIIKL 203
DB 210 RVIQM 214

RESULT 13

US-09-815-242-5642

; Sequence 5642, Application US/09815242

; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5642

; LENGTH: 249

; TYPE: PR1

; ORGANISM: Staphylococcus aureus

US-09-815-242-5642

Query Match 33.0%; Score 336; DB 9; Length 249;

Best Local Similarity 34.7%; Pred. No. 2.3e-27;

Matches 74; Conservative 47; Mismatches 82; Indels 10; Gaps 3;

QY 1 MIEINDLKSPGVRILWQGL---SHKPLPGTMTALTGASGSGKSTLLNCLGTLDPKSSG 57

DB 3 ILSVQHVSKTYKQKHTFQALKDINFQKGEFVALMGPSGSGKTLNVLSSIDQISSGS 62

QY 58 ILVEDVLLKSTRKQRYKNTVGYLFQDVALIPDRTVKFNQLAVEKHKWPFIQ--- 114

DB 63 VIANGQELNKLNQALAKFRKESLGTFIDYISLFTLVKENIMLPLSVQKMSKATMEKN 122

QY 115 ---VLHAGVLESPEEKVPFELSGGEQORTALARVLLKNPRIILADEPTGALDITNSSELVI 171

DB 123 YKAITTALGIDYLNKYPSELSSGQOQRTAAARAFVHKPQIIFADEPTGALDITNSSELVI 182

QY 172 EALRALAD-KATVVVATHSPRESADTIKL 203

DB 183 QRLEEMNKSFDITIVMVTHDFTAAASFAERVIML 215

RESULT 14

US-09-815-242-12706

; Sequence 12706, Application US/09815242

; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert


```
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/205,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12706
; LENGTH: 252
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-12706

Query Match      33.0%; Score 336; DB 9; Length 252;
Best Local Similarity 34.7%; Pred.No.2.3e-27;
Matches 74; Conservative 47; Mismatches 82; Indels 10; Gaps 3;

QY 1 MIEINDLKXSGFVRILWQGL---SHKELPGTWTALTGASGSGKSTLLNCLGTLDPKPSGQ 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 ILSVQHVSKTYGKXHTFQALKDNFDIQKGEFVAIMGPSGKTTLLNVSSIDQISSGS 62

QY 58 ILVEDVDLLKLSRKQRLRYNTVGYLFQDYALIPDRTVKFNQLAVEKHKEWPEIQ--- 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 VIANGQELNKLNQAKAFKRESIGFTFDYSILPTLVKENIMPLSLVQKMSKATMEKN 122

QY 115 ---VLHAVGLSPPEKPVFELSGGEQQTALARVLLKNPRIILADEPTGALDITNSELVI 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 YKAITTALGIYDLGNKYPSELSGGQQQTAAARAFVHKPQIIFADEPTGALDSKANDLL 182

QY 172 EALRALADK-GATVVVATHSPLPRESADTIKL 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 QRLEENKNSFDITIVMVTHDPVAASFAERVIML 215
```

```
RESULT 15
US-10-238-075-1256
; Sequence 1256, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1256
; LENGTH: 231
; TYPE: PR1
```

```
; ORGANISM: Escherichia coli
US-10-238-075-1256

Query Match      32.7%; Score 333.5; DB 14; Length 231;
Best Local Similarity 38.0%; Pred.No.3.6e-27;
Matches 78; Conservative 32; Mismatches 92; Indels 3; Gaps 1;

QY 2 IEINDLKXSGFVRILWQGLSHKFLPGTWTALTGASGSGKSTLLNCLGTLDPKPSGQILVE 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 IETRHLYKRFQDVTDALDDINLRIAQCEFAVIMGASGSGKTTLLMNITICLTATGQVFLD 70

QY 62 DVDLLKLSRKQRLRYNTVGYLFQDYALIPDRTVKFNQLAVEKHKEWPE---IPQVLHA 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 GTDAALDEEGRFRFAEKIGLVFQOHLIPTLTALENIMLAQHYHSVVDEAARAKVLEQ 130

QY 119 VGLESFEEKPVFELSGGEQQTALARVLLKNPRIILADEPTGALDITNSELVIEALRALA 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 VGLGHRVTHLPQLSGGEQQRVCARALVNEPPIFADEPTGNLDEENEQRLDLDLTLH 190

QY 179 DKGATVVVATHSPLPRESADTIKL 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 ROGRITIVMVTHNPALGQFADRIILRL 215

Search completed: July 27, 2004, 16:41:55
Job time : 47 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:33:10 ; Search time 16 seconds
(without alignments)
1220.430 Million cell updates/sec

Title: US-09-868-338-9
Perfect score: 1019
Sequence: 1 MIEINDLKKSFGVRLWQGL.....VVVATHSPFPRESADTIKL 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*

1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	408	40.0	213	2 G95079	ABC transporter, A
2	405	39.7	213	2 B97947	hypothetical prote
3	399	39.2	213	2 E95232	ABC transporter, A
4	399	39.2	213	2 G98096	hypothetical prote
5	386	37.9	213	2 A95013	hypothetical prote
6	385	37.8	213	2 D97884	hypothetical prote
7	380.5	37.3	210	2 G95228	ABC transporter, A
8	380.5	37.3	210	2 C98093	hypothetical prote
9	375	36.8	228	2 A84088	ABC transporter (A
10	373	36.6	211	2 B86626	ABC transporter AT
11	362.5	35.6	465	2 S73400	glutamine transpor
12	361.5	35.5	226	2 D69433	ABC transporter, A
13	358	35.1	331	2 AB3297	ABC transporter AT
14	355	34.8	213	2 A83585	ABC transporter (A
15	354.5	34.8	228	2 B69377	ABC transporter, A
16	353	34.6	227	2 A84109	ABC transporter (A
17	353	34.6	466	2 B64207	heterocyst maturat
18	351	34.4	232	2 C97089	ABC transporter, A
19	350.5	34.4	300	2 A82932	ABC transporter UU
20	349.5	34.3	255	2 B97087	ABC-type transport
21	349	34.2	246	2 E90216	ABC transporter, A
22	347	34.1	648	1 G64826	probable ABC trans
23	346	34.0	231	2 B69477	ABC transporter, A
24	345.5	33.9	217	2 D81263	probable ABC trans
25	344	33.8	207	2 A81421	ABC transporter, A
26	344	33.8	648	2 E90749	hypothetical prote
27	344	33.8	648	2 A85600	hypothetical prote
28	342.5	33.6	227	2 F96955	ABC-transporter, A
29	342.5	33.6	230	2 T34622	ABC-transporter AT

Query Match 40.0%; Score 408; DB 2; Length 213;

Best Local Similarity 41.9%; Pred. No. 4.4e-28;

Matches 88; Conservative 44; Mismatches 58; Indels 10; Gaps 3;

QY 1 MIEINDLKKSFGVRLWQGLSHKFLPGTWTALTGASGSGKSTLNCILGTLDKPSSGQILV 60

DB 1 MIEINKSKFSSQLPSDMNLHFEGKIVALTIGSGGKTLTLLNMIGRLPYDKQILIY 60

QY 61 EDVLLKLSTRKQRLYRNTVGYLFQDYALIPDRTVFNQLAV-----EKHKWPEIPQ 114

DB 61 DGTSLKDI---KPSVFFRDYGLFQDFGLIESQTVKENLNLGLVKKLKEKISLMKQ 117

QY 115 VLHVGLESFE-EKPVFELSGEQRTALARVLKKNPRIILADEPTGALDITNSELVIEA 173

DB 118 ALNRVNLGYLDLKKQIFELSGGEAQRVALAKIILKDPPLILADEPTASLPKXSEELLSI 177

QY 174 LRALADKATVVVATHSPFPRESADTIKL 203

DB 178 LESLKNPRTIIATHNPLIWEQVDQVIRV 207

RESULT 2

B97947

hypothetical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: B97947

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgess, S.; DeHoff, B.S.; Es
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.

30 342 33.6 207 2 AC1796 ABC transporter, A
31 342 33.6 304 2 F95337 probable ABC trans
32 341.5 33.5 237 2 T47012 hypothetical prote
33 341.5 33.5 237 2 AE0237 ABC transporter, A
34 340 33.4 233 2 A11580 ABC transporter, A
35 339 33.3 259 2 F86843 ABC transporter AT
36 339 33.3 648 2 AF0609 conserved hypothet
37 338.5 33.2 666 2 AHI167 ABC transporter, A
38 337 33.1 252 2 D90079 ABC transporter prote
39 337 33.1 257 2 A70074 ABC transporter (A
40 336 33.0 228 2 H72342 ABC transporter, A
41 336 33.0 236 2 H98058 hypothetical prote
42 334.5 32.8 238 2 G96929 ABC transporter AT
43 333 32.7 233 2 AGL227 ABC transporter, A
44 333 32.7 233 2 C95192 ABC transporter, A
45 331 32.5 234 2 C97001 ABC transporter, A

ALIGNMENTS

RESULT 1

G95079

ABC transporter, ATP-binding protein SP0687 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: G95079

R:Tettelin, H.; Nelson, X.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: G95079

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-213 <CUR>

A:Cross-references: GB:AB005672; PIDN:AAK74832.1; PID:G14572163; GSPDB:GN00164; TIGR:SP45

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0687

J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E97947
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99406.1; PID:g15458184; GSPDB:GN00174
C:Genetics:
A:Gene: ABC-NBD

Query Match 39.7%; Score 405; DB 2; Length 213;
Best Local Similarity 41.9%; Pred. No. 7.9e-28;
Matches 88; Conservative 43; Mismatches 69; Indels 10; Gaps 3;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60
DB 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60
QY 61 EDVDLLKLSRQRLYRKNVTGYLFQDYALIPDRVTFRNQLAV-----EKHKWPEIQ 114
DB 61 DGTSLKDI---KPSVFRDYLGYLFQDFGLIESQTVKENLNLGVLGKLEKEKISLMKQ 117
QY 115 VLHVGLESFB-EKPVFELSGEQORTALARVLKNPRIILADEPTGALDITNSLVIEA 173
DB 118 ALNRVNLISYLDLQKPIFELSGEQORTALARVLKNPRIILADEPTGALDITNSLVIEA 177
QY 174 LRALADKATVVVATHSPRESADTIKL 203
DB 178 LESLKNPRTIIIIATHNPLVWKNKADEIIDM 207

RESULT 3
E95232
ABC transporter, ATP-binding protein SP1987 (imported) - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: E95232
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: E95232
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK76054.1; PID:g14973495; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1987

Query Match 39.2%; Score 399; DB 2; Length 213;
Best Local Similarity 42.9%; Pred. No. 2.6e-27;
Matches 90; Conservative 40; Mismatches 70; Indels 10; Gaps 3;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60
DB 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60
QY 61 EDVDLLKLSRQRLYRKNVTGYLFQDYALIPDRVTFRNQLAV-----EKHKWPEIQ 120
DB 61 QKDKLTIPTRE---YFRDQGYLFQDFGLIENSIKENLDLGFVGQKISKVERLERQVG 117
QY 121 -LESFE-----EKPVFELSGEQORTALARVLKNPRIILADEPTGALDITNSLVIEA 173
DB 118 ALEKVNGLVDLEQKIYTLGSGEQORTALARVLKNPRIILADEPTGALDITNSLVIEA 177
QY 174 LRALADKATVVVATHSPRESADTIKL 203
DB 178 LESLKNPRTIIIIATHNPLVWKNKADEIIDM 207

Query Match 37.9%; Score 386; DB 2; Length 213;
Best Local Similarity 40.2%; Pred. No. 3.6e-26;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60
DB 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60

DB 178 LVLDKDNRIIIIIATHNPLVWKNKADEIIDM 207

RESULT 4
G98096
hypothetical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: G98096
R:Hoskins, J.A.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Mc
e, R.; LeBlanc, D.J.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Es
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G98096
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK00604.1; PID:g15459486; GSPDB:GN00174
C:Genetics:
A:Gene: ABC-NBD

Query Match 39.2%; Score 399; DB 2; Length 213;
Best Local Similarity 42.9%; Pred. No. 2.6e-27;
Matches 90; Conservative 40; Mismatches 70; Indels 10; Gaps 3;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60
DB 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60
QY 61 EDVDLLKLSRQRLYRKNVTGYLFQDYALIPDRVTFRNQLAV-----EKHKWPEIQ 120
DB 61 QKDKLTIPTRE---YFRDQGYLFQDFGLIENSIKENLDLGFVGQKISKVERLERQVG 117
QY 121 -LESFE-----EKPVFELSGEQORTALARVLKNPRIILADEPTGALDITNSLVIEA 173
DB 118 ALEKVNGLVDLEQKIYTLGSGEQORTALARVLKNPRIILADEPTGALDITNSLVIEA 177
QY 174 LRALADKATVVVATHSPRESADTIKL 203
DB 178 LVLDKDNRIIIIIATHNPLVWKNKADEIIDM 207

RESULT 5
A95013
hypothetical protein SP0111 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: A95013
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: A95013
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74298.1; PID:g14971578; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0111

Query Match 37.9%; Score 386; DB 2; Length 213;
Best Local Similarity 40.2%; Pred. No. 3.6e-26;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60
DB 1 MIEINDLKXFGVIRILWQGLSHKFLPGTMTALTGASGSKSTILNCLGTLDPKSSGQILV 60

[illegible]

RESULT 6

D37894
 Hypochemical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C:Accession: D37894
 C:Authors: J.A.; Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgess, S.; DeHoff, B.S.; E
 e, R.; LeBlanc, D.J.; Les, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 Y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; J
 A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21423245; PMID:11544234
 A:Accession: D37894
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-213 <KUR>
 A:Cross-references: GS:AE007317; PIDN:AAK98904.1; PID:g15457636; GSPDB:GN00174
 C:Genetics:
 A:Gene: ABC-NBD

	Query Match	37.8%;	Score 385;	DB 2;	Length 213;
	Best Local Similarity	39.7%;	Pred. No. 4.3e-26;		
	Matches 83; Conservative	46; Mismatches 74;	Indels 6;	Gaps 1;	
Cy	1	MIENDLKXGFGVRIILWQGLSHKPLPCTMTALTGASGSCKTLNCLGTLDKPSSGGIYL	60		
Dd	1	MIELKNITKTGGVLNDLSLRDQDLVAIVGSGSKSTLLNLGLIDGDYSRGYEI	60		
Cy	61	EDVDLLKLSTRQRLYRKYLTGVGYLFQDYALIFDRTVFNIQLAV-----EKHKWEPIQ	114		
Dd	61	FQGNTLAVNSAKSQTIIREHISYLFQNFALIDDTEVNLMALKYVKLPKKOLKKVEE	120		
Cy	115	VLVHAVGLESEEEKVPFSLSCGEQORTALARVLLKNPRIIIDAEPTGALDI TNSELVTIAL	174		
Dd	121	I LERVGUSATHQRVSLSGGEQQOIRAVARAILPSOLI LADEPTGSLDPENRDVLVKFL	180		
Cy	175	RALADKGATVVVATHSPFLFRESADTI IKL	203		
Dd	181	LEMNRGKTVTIVTHDAVAAOCCHRVTEV	209		

RESULT 7

ABC transporter, ATP-binding protein SP1957 [imported] - Streptococcus pneumoniae (strain G95228)
 C;Species: Streptococcus pneumoniae
 C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
 C;Accession: G95228
 R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
 nson, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A.; Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
 A;Reference number: A95000; MUID:21357209; PMID:11463916
 A;Accession: G95228
 A;Status: preliminary
 A;Molecule type: DNA

A;Residues: 1-210 <KUR>
A;Cross-references: GB:
A;Experimental source:
C;Genetics:
A;Gene: SP1957

[illegible]

RESULT 8

C98093
hypothetical protein ABC-NBD [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: C98093
e,Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff,
R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAB,
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskun
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A57872; MUID:21423245; PMID:11544234
A:Accession: C98093
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-210 <R>
A:Cross-references: GB:AE007317; PID:1545459455; GSPDB:GN00174
C:Genetics:
A:Gene: ABC-NBD

Query Match 37.3%; Score 380.5; DB 2; Length 210;
Best Local Similarity 41.0%; Pred. No. 1e-25;
Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4.

	QY	1	MIEINDUKSGFVIRLWQG:SHKPLPOTMTALTGASGSKTLLNGLCTIDTKFPSSQILV 60
	Dd	1	MIELQVSKSFGERELFSNLMTFEAGKVVALIGSSGKTTLMNMIGKLE-PYDGTIFY 59
	QY	61	EDVLLKLSTRKQLRYEKVTGYLFQDYALIPDRTVKFNIQLAVEKHKEPEI-----PQ 114
	Dd	60	RKZD---LANYSDDPRHELGVLPFNQGLIENQS:EENKLGLTGQKLSRSEQRLRQK 116
	QY	115	VLHVAQGESFE-BKYPFELSGGQQOQTALARULLKNPRIILADEPTGALDLINSELVIEA 173
	Dd	117	ALEQVGLVYLDLCKRFELSGGSORVALAKIILKNPPFIILADEPTASIDPATSQLIMEI 176
	QY	174	LRLADKGATVVVATHSPFLRESADTIIKL 203
	Dd	177	LLSRDDNRLLIIITATNPAINWEADVEFTM 206

RESULT 9

R50011 5
A84088 - Bacillus halodurans (strain C-
C) Species: Bacillus halodurans [imported] BH3505

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 17-May-2002
C;Accession: A84088
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: A84088
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-228 <STO>
A;Cross-references: GB:AP001519; GB:BA000004; NID:G10176109; PIDN:BA07224.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
C;Superfamily: short-chain ATP-binding cassette proteins; ATP-binding cassette homology
Query Match 36.8%; Score 375; DB 2; Length 228;
Best Local Similarity 38.1%; Pred. No. 3.5e-25;
Matches 82; Conservative 46; Mismatches 75; Indels 12; Gaps 3;
Qy 1 MIBINDLKSPGVRLWQGLSHKFL---PGTWTALTGASGSGKSTLNCGLTLDKPSG 56
Dy 1 MIQEGITFTFNNAVTFQALANVSLIKESGFVAITGTSGSKTLLNGLCLDDPTSG 60
Qy 57 QILVEDVLLKSTRKQRLYKNTVGYLPQDYALIPDRVTKFNQLAV-----EKHK 108
Dy 61 KYSLTGEDVSQYNKTKKALLRNMIGFVLQDFALVHYTVKQNVMLPLMYVGNKKKKER 120
Qy 109 WPIPIQVHLAVGLESPFVPELSSGSGQRTALRVLKPRILADEPTGALDNLNSEL 168
Dy 121 QKXIEBLLXGIAHKEKERTALLSGGQKQRTAIGRALINEPRILADEPTGALDQKTSK 180
Qy 169 LVTEALRALADKATVVVATHSPFRESADTIKL 203
Dy 181 EIMELNKLHKEGKTVIIIVTHDPFVADYCDRIQL 215
RESULT 10
B86626
ABC transporter ATP-binding protein yabE [imported] - Lactococcus lactis subsp. lactis
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: B86626
R;Bolotin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: B86626
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-211 <STO>
A;Cross-references: GB:AE005176; PID:G12722846; PIDN:AAK04108.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yabE
Query Match 36.6%; Score 373; DB 2; Length 211;
Best Local Similarity 39.2%; Pred. No. 4.7e-25;
Matches 82; Conservative 42; Mismatches 79; Indels 6; Gaps 1;
Qy 1 MIBINDLKSPGVRLWQGLSHKFLPGTWTALTGASGSGKSTLNCGLTLDKPSGQILV 60
Dy 1 MIBIEBLSYKGIHIFDKNLRIPEGKMTAIGTSGGAKSTLLNIGLIEDYDDGKYIF 60
Qy 61 EDVDLLKSTRKQRLYKNTVGYLPQDYALIPDRVTKFNQLAV-----EKHKWPEIPQ 114
Dy 61 NGQAFPPNSSLAKMKRNKISLYLFQNFALLEDEIERNLIALIYSRISKKEKMKK 120
Qy 115 VLHVGLESPFVPELSSGSGQRTALRVLKPRILADEPTGALDNLNSELVIAL 174
Dy 121 LLQGVINHLNATKVYSLSGGKQRTAIGRALINEPRILADEPTGSLDTENRNREVIALL 180
Qy 175 RALADKATVVVATHSPFRESADTIKL 203

Db 181 RQEVDKGRAVIVTHDSVLKESVDLVEI 209
RESULT 11
S73400
glutamine transport ATP-binding protein glnQ - Mycoplasma pneumoniae (strain ATCC 29342)
N;Alternate names: hypothetical protein R02_orf465
C;Species: Mycoplasma pneumoniae
A;Variety: ATCC 29342
C;Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 02-Feb-2001
C;Accession: S73400
R;Himmelfreid, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.
A;Reference number: S73327; MUID:97105885; PMID:8948633
A;Accession: S73400
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-465 <HIM>
A;Cross-references: EMBL:AE000009; GB:U00089; NID:G1673720; PIDN:AAE95722.1; PID:G1673720
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Gene: glnQ
A;Genetic code: SGC3
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop
F;251-445/Domain: ATP-binding cassette homology <ABC>
F;268-275/Region: nucleotide-binding motif A (P-loop)
Query Match 35.6%; Score 362.5; DB 2; Length 465;
Best Local Similarity 36.7%; Pred. No. 1.1e-23;
Matches 77; Conservative 54; Mismatches 70; Indels 9; Gaps 4;
Qy 1 MLEINDLKK--SFGV--RILWGLSHKFLPGTWTALTGASGSGKSTLNCGLTLDKPSG 56
Dy 231 ILEKNVYKTYNGVTTNVALKGDILKKAHDFVILGPSGSKTLLNLSGMDRPSG 290
Qy 57 QILVEDVLLKSTRKQRLYKNTVGYLPQDYALIPDRVTKFNQLAVKHKWPE---I 112
Dy 291 SVVANGQEMICMNDRLTNFRNRYGVYFQYGLLPNLTVRENVGVANLQRPDKRNI 350
Qy 113 POVLHVGLESPFVPELSSGSGQRTALRVLKPRILADEPTGALDNLNSELVIE 172
Dy 351 DELLAVGKHLQKLPNELSGGQQRVSTARAKNPLIFGDEPTGALDLEWTQIVLK 410
Qy 173 ALRALADK-GATVVVATHSPFRESADTI 201
Dy 411 QFLAKQRYKTYMTVITVTHNLIQLADLVI 440
RESULT 12
D69433
ABC transporter, ATP-binding protein homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001
C;Accession: D69433
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,
R.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kinkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Uetebach, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: D69433
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-226 <KLE>
A;Cross-references: GB:AE001001; GB:AE000782; NID:G2689324; PIDN:AAE89781.1; PID:G2649101
C;Superfamily: short-chain ATP-binding cassette proteins; ATP-binding cassette homology
C;Keywords: ATP; nucleotide binding; P-loop
F;23-219/Domain: ATP-binding cassette homology <ABC>

NUCLEAR ACTORS REG. NO. 4317-4331, 2000

170 VIEALRALAD-KGATVVVATHSPRESADTIKL 203

us-09-868-338-9.rpr

Mon Aug 2 09:36:56 2004

Db 182 VMEIRHCNEVLGVTIILVTHDPSLAKYGDRIKL 216

Search completed: July 27, 2004, 16:36:52
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:29:29 ; Search time 13 Seconds

(without alignments)
813.095 Million cell updates/sec

Title: US-09-868-338-9

Perfect score: 1019

Sequence: 1 MEINDLKKSFGVRIILWQSL.....VVVATHSPILFRSADTIITKL 203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	362.5	35.6	465	1 Y065 MYCPN	P75612 mycoplasma
2	353	34.6	466	1 Y065 MYCGE	P47311 mycoplasma
3	347	34.1	648	1 MACB ECOLI	P75831 escherichia
4	337	33.1	257	1 YXDL BACSU	P42423 bacillus su
5	329.5	32.3	248	1 Y986 MYCTU	O53899 mycobacteri
6	325.5	31.9	284	1 Y46B MYCGE	O92670 mycoplasma
7	325	31.9	240	1 Y01Z BACSU	P54537 bacillus su
8	323	31.7	234	1 Y352 THEMA	O9WY17 thermotoga
9	312.5	30.7	284	1 Y46B MYCPN	O50316 mycoplasma
10	311.5	30.6	235	1 Y796 METJA	O58206 methanococc
11	309.5	30.4	233	1 LOLE ECOLI	P75957 escherichia
12	309	30.3	224	1 YF08 METJA	O58903 methanococc
13	309	30.3	344	1 METN VIBCH	O9Ktj5 vibrio chol
14	303	29.7	247	1 YXKI BACSU	P39456 bacillus su
15	302.5	29.7	249	1 YXEO BACSU	P54954 bacillus su
16	301	29.5	330	1 YP64 MYCTU	O50734 mycobacteri
17	300	29.4	377	1 POTG ECOLI	P31134 escherichia
18	296	29.0	250	1 YECC ECOLI	P37774 escherichia
19	295	28.9	345	1 METN HAEIN	P44785 haemophilus
20	291	28.6	222	1 FTSE ECOLI	P10115 escherichia
21	289.5	28.4	231	1 YTRE BACSU	O34392 bacillus su
22	289	28.4	343	1 METN YERPE	O82h38 yersinia pe
23	287.5	28.2	221	1 Y700 RICPR	O92cm4 rickettsia
24	287.5	28.2	229	1 LOLE BUCAP	O44613 buchneria ap
25	286	28.1	231	1 LOLE NEIMP	P57031 neisseria m
26	285	28.0	343	1 METN ECOL6	O8X729 escherichia
27	285	28.0	343	1 METN ECOLI	P30750 escherichia
28	282.5	27.7	231	1 LOLE NEIMA	P57030 neisseria m
29	282.5	27.7	348	1 CYSA_XLTP1	O87dc9 xyloella fas
30	282	27.7	355	1 CYSA_XLTP3	P74503 synecocyst
31	280.5	27.5	339	1 Y467 MYCPN	P75110 mycoplasma
32	280	27.5	242	1 GLNQ BACST	P27675 bacillus st
33	279.5	27.4	227	1 LOLE_HAEIN	P45247 haemophilus

ALIGNMENTS

RESULT 1

ID	Y065 MYCPN	STANDARD;	PRT;	465 AA.
AC	P75612;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical ABC transporter ATP-binding protein MG065 homolog (R02 orf465).			
DE	(R02 orf465).			
GN	MPN081 OR MP074.			
OS	Mycoplasma pneumoniae.			
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=2104;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 29342 / M129;			
RX	MEDLINE=97105885; PubMed=8948633;			
RA	Himmelreich R., Hilbert H., Pragens H., Pirkl E., Li B.-C., Herrmann R.;			
RA	"Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae."			
RT	Nucleic Acids Res. 24:4420-4449(1996).			
CC	-1- SIMILARITY: Belongs to the ABC transporter family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AB000009; AAB95722.1; -			
DR	PIR; S73400; S73400.			
DR	InterPro; IPR003593; AAA_ATPase.			
DR	InterPro; IPR003439; ABC_transporter.			
DR	Pfam; PF00005; ABC_tran; 1.			
DR	ProDom; PD000006; ABC_transporter; 1.			
DR	SMART; SM00382; AAA; 1.			
DR	PROSITE; PS00211; ABC_TRANSPORTER 1; 1.			
DR	PROSITE; PS00893; ABC_TRANSPORTER 2; 1.			
KW	Hypothetical protein; ATP-binding; Transport; Complete proteome.			
FT	NP_BIND 268 275 ATP (POTENTIAL).			
SQ	SEQUENCE 465 AA; 53511 MW; P8F85014C4A1A25D CRC64;			

Query Match 35.6%; Score 362.5; DB 1; Length 465;				
Best Local Similarity 36.7%; Pred. No. 3.1e-23;				
Matches 77; Conservative 54; Mismatches 70; Indels 9; Gaps 4;				
QY	1	MEINDLK--SFGV--RILWQGLSHKFLPCTMTALTGASGSGKSTLNCIGLTKDPSSG	56	
Db	231	IEUKNWKYITNGVTNAVLRKGLDLKLAHDFVILGPSGSGKATLLNLSIGMDPSSG	290	
QY	57	QILVEDVLLKLSTRKQRLKRNVTGYLFQDYALIPDRTVKFNQLQAVEKHKPE---	112	
Db	291	SVVWNGQEMICMNDQLNTFRNRYVGVYFQYGLPLNLTVRNVGVGANLQRNPDKRNI	350	


```
DR EMBL; AB071146; BAB64542.1; -
DR EMBL; AB000189; AAC73966.1; -
DR EMBL; D90725; BAA35598.1; -
DR PIR; G64826; G64826.
DR EcoGene; EG13695; macB.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR003838; DUF214.
DR Pfam; PF00005; ABC_tran; 1.
DR Pfam; PF02687; FtsX; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport; Transmembrane; Inner membrane;
KW Antibiotic resistance; Complete proteome.
FT NP_BIND 41 48 ATP (POTENTIAL).
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 417 437 POTENTIAL.
FT TRANSMEM 523 543 POTENTIAL.
FT TRANSMEM 576 596 POTENTIAL.
FT TRANSMEM 612 632 POTENTIAL.
SQ SEQUENCE 648 AA; 70702 MW; 157B7D61ED17346 CRC64;
Query Match 34.1%; Score 347; DB 1; Length 648;
Best Local Similarity 36.9%; Pred. No. 9.4e-22;
Matches 79; Conservative 42; Mismatches 81; Indels 12; Gaps 3;
QY 1 MIEINDLKXSF-GRILMQLSHKFLPGTWTALTGASGSKSTLNLCLGTLDPKPS 55
DB 4 LLELKDIRSYRYPAGDEQVVL-KGISLDIYAGENVAIVGASGSKSTLNLCLGTLDPKPS 62
QY 56 GQILVEDVLLKSTKQRLYKNTVGYLFQDYVALIPDRTVKFNQLAV-----EKHW 109
DB 63 GTRVAGQDVATLADALAQIRHFGIFQRVHLSHLTAEQNVFVAYGLERKQRL 122
QY 110 PEIPQVILHVGLESFEKPFVLSGSGEQRTALARVLLKNPRLADEPTGALDLTNSEL 169
DB 123 LRAQELLQRLGLEDRTEVYPAQLSGGQQRVSARALMNGGVILADEPTGALDSHSEE 182
QY 170 VIEALRALADKATVTVATHSPFRESADTIKL 203
DB 183 VMAILHOLDRGHTVIVTHDPQVAAQAEVIEI 216
RESULT 4
YXDL_BACSU STANDARD; PRT; 257 AA.
AC P42423;
DT 01-NOV-1995 (Rel. 32, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Hypothetical ABC transporter ATP-binding protein yxdl.
GN YXDL OR B65F OR BSU39640.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC141;
RX MEDLINE=95039891; PubMed=7952181;
RA Yoshida K., Sano H., Miwa Y., Ogasawara N., Fujita Y.;
RT "Cloning and nucleotide sequencing of a 15 kb region of the Bacillus
RT subtilis genome containing the iol operon."
RL Microbiology 140:2289-2298(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC141;
RX MEDLINE=97021444; PubMed=8867804;
RA Yoshida K.-i., Fujimura M., Yanai N., Fujita Y.;
RT "Cloning and sequencing of a 23-kb region of the Bacillus subtilis
RT genome between the iol and hut operons."
RL DNA Res. 2:295-301(1995).
```

```
RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Allioni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherst S.,
RA Borrius R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brulliet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Huilo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.F., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Plescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche M., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Darchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. STRONG, TO
CC E.COLI FTSE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; D14399; BAA03302.1; -
DR EMBL; D45912; BAA08315.1; -
DR EMBL; Z99124; CAB16000.1; -
DR PIR; A70074; A70074.
DR Subtilist; BG11128; yxdl.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; ATP-binding; Transport; Complete proteome.
FT NP_BIND 40 47 ATP (POTENTIAL)
SQ SEQUENCE 257 AA; 28722 MW; D9CE0LE9ECCIA6AD CRC64;
Query Match 33.1%; Score 337; DB 1; Length 257;
Best Local Similarity 37.3%; Pred. No. 2.2e-21;
Matches 76; Conservative 44; Mismatches 70; Indels 14; Gaps 5;
QY 1 MIEINDLKXSFGRILLWGL---SHKFLPGTWTALTGASGSKSTLNLCLGTLDPKPSGQ 57
DB 4 MLEVHKNKTYKGVSYQALQKQISFIEEGFTAVMGSGSKTKTLNIIITIDRPSGD 63
QY 58 ILVEDVLLKSTKQRLYKNTVGYLFQDYVALIPDRTVKFNQLAVKHWPIPOV-- 115
DB 64 ILINGENPHRLKRTKLAHFRKRLGFGVFDENLLDTLTIGENIMPLTLEK--EAPSWE 121
```

	FT	NP_BIND	43	50	ATP (POTENTIAL).
	SQ	SEQUENCE	248 AA;	27373 MW;	71LEUBCB872EB8.A CRC64;
		Query Match	32.3%;	Score 329.5;	DB 1; Length 248;
		Best Local Similarity	35.5%;	Pred. No. 9.1e-21;	
	Matches	76; Conservative	47; Mismatches	80; Indels	11; Gaps 4;
QY	1	MEINDUKKSPGV-----RIWQLGSLHKFLPGTWTALTGCASGSKSTLLNCLGTLDKPSSG	56		
Db	6	IVQLSNLSWTTFREGETRRQVLDHITDFEFGEFVALLGQSGLKTLNLNISIGIERPTTG	65		
QY	57	QILVEDVDLLKLSTRKORLYRKNTVGVLFDYALIPDRTVKFNLQL-----AVEKHWPKE	111		
Db	66	DVTINGFAITOKTERDRTLFRDOIGIVFOFFNNLTPLTVLENITLPQELAGVSQRKA	125		
QY	112	IPO-VLVHAVGLSEFEPEKFVELSGEQBOORTALARVLKNPRILLADEPTCALDLTNSELV	170		
Db	126	VARDILLEKVGNADRETRFPDKLSGSEQRVASELAHNPNMLVADEPTGNLSDTDGDKV	195		
QY	171	IEALRALA-DKGATVVVATHSPLPRESADTIKL	203		
Db	186	LVDLDTLRQAQKTLIIVATHSPSTMQTADRNVNL	219		
 RESULT 6 Y46B MYCGE					
ID	Y46B MYCGE	STANDARD;	PRT;	284 AA.	
AC	Q9ZBT70;				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Hypothetical ABC transporter ATP-binding protein MG468.1.				
DN	MG468.1.				
OS	Mycoplasma genitalium.				
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.				
OX	NCBI_TaxID=2097;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 33530 / G-37;				
RX	MEDLINE=96026346; PubMed=7569993;				
RA	Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,				
RA	Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,				
RA	Fritschman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,				
RA	Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,				
RA	Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,				
RA	Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;				
RT	"The minimal gene complement of Mycoplasma genitalium.";				
RL	Science 270:397-403(1995).				
RN	[2]				
RP	IDENTIFICATION.				
RA	Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,				
RA	Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,				
RA	Fritschman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,				
RA	Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,				
RA	Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,				
RA	Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;				
RL	Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.				
CC	-!- SIMILARITY: Belongs to the ABC transporter family.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	EMBL; U39728; AAC72489.1; -.				
CC	TIGR; MG468.1; -.				
DR	InterPro; IPR003593; AAA_ATPase.				
DR	InterPro; IPR003439; ABC_transporter.				
DR	Pfam; PF00005; ABC_tran_1				
DR	ProDom; PD000006; ABC_transporter; 1.				
KW	Protein; Complete proteome.				

```
DR SMART: SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; ATP-binding; Transport; Complete proteome.
FT NP_BIND 89 ATP (POTENTIAL).
SQ SEQUENCE 284 AA; 32256 MW; A6967485FBA43177 CRC64;

Query Match 31.9%; Score 325.5; DB 1; Length 284;
Best Local Similarity 36.3%; Pred. No. 2.3e-20;
Matches 70; Conservative 44; Mismatches 76; Indels 3; Gaps 2;

QY 14 RLWQGLSHKFLPDTMTALTGASGSKSTLLNCLGTLDPKSSQOILVEDLLKLSRKQ 73
DQ :::::
DB 69 QLQCNHINKLPGFVVLGSGSKTSLLSLSALDRPTSGDSFVCGNTTCCSDAKL 128
QY 74 RLYRKNVTGYLFQDYALIPRTVKFNQLA--VEKHKWPPIQVILHVGLESEEEKPVSE 131
DQ :::::
DB 129 TALRNKNVGFYFQYGLRDLVDNKLALPKKFNNNLELLELLEKHEHKKVHK 188
QY 132 LSGEGQORTALARKVLLKNPRIILADEPTGALDITNSELVIE-ALRALADKATVVVATHS 190
DQ :::::
DB 189 LSGGQQQRVALARALIKEPKILFGDEPTGAVNIDSKKILQFFVEYNRDAGTIVIVTEN 248
QY 191 PLFRESADTIKL 203
DQ :::::
DB 249 EKIVELAKRVIKI 261

RESULT 7
YQIZ_BACSV STANDARD; PRT; 240 AA.
AC PS4537;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable amino-acid ABC transporter ATP-binding protein yqiz.
GN YQIZ OR BSU32960.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many
RT speculation genes."
RL Microbiology 142:3103-3111(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferzari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Huilo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Fucic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
```

```

[1]
SEQUENCE FROM N.A.
STRAIN-ATCC 29342 / M129;
MEDLINE=96177562; PubMed=8604303;
Hilbert H., Himmelreich R., Plagens H., Herrmann R.;
"Sequence analysis of 56 kb from the genome of the bacterium
Mycoplasma pneumoniae comprising the dnaA region, the atp operon and
a cluster of ribosomal protein genes.";
Nucleic Acids Res. 24:628-639(1996).
[2]
SEQUENCE FROM N.A.
STRAIN-ATCC 29342 / M129;
MEDLINE=97105885; PubMed=89486633;
Himmelreich R., Hilbert H., Plagens H., Pirkle E., Li B.-C.,
Herrmann R.;
"Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae.";
Nucleic Acids Res. 24:4420-4449(1996).
-!- SIMILARITY: Belongs to the ABC transporter family.
-
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch)
-
EMBL; U34816; AAC43649.1; -.
EMBL; AE000017; AAB95805.1; -.
PIR; S62839; S62839.
InterPro; IPR003593; AAA ATPase.
InterPro; IPR003439; ABC transporter.
Pfam; PF00005; ABC tran; 1.
ProDom; PD000006; ABC transporter; 1.
SMART; SM00382; AAA; 1.
PROSITE; PS00211; ABC TRANSPORTER 1; 1.
PROSITE; PS00893; ABC TRANSPORTER 2; 1.
Kw Hypothetical protein; ATP-binding; Transport; Complete proteome.
FT NP_BIND 89 96 ATP (POTENTIAL).
SQ SEQUENCE 284 AA; 32241 MW; E2275C81CFE295E0 CRC64;

Query Match 30.7%; Score 312.5; DB 1; Length 284;
Best Local Similarity 37.0%; Pred. No. 2.9e-19;
Matches 71; Conservative 36; Mismatches 66; Indels 19; Gaps 4;

QY 23 KFLPTMTALTGASGSGKSTLLNCILGTLDKPSGG-----QILVEDVDLLKLSRKQR 74
DB 78 KIAPEFVVIIGKSGSGKTSLLSLADRTSGVSFVCGRSTTCNDAQLTSL----- 131
QY 75 LYRNTVGYLFQDYALIPIRTVKFNQLQAV--ERHKWPEIPOVLHAVGLSEFEEKPVPEL 132
DB 132 --RNNVGYIFQYGLLEDLVNDONIKLAVFPFKRHHNNLLELLERLELKEHRNKKITKL 189
QY 133 SCGEQRTALRVLLKNPRIILADEPTGALDITNSLVIE-ALRALADKATVVVATHSP 191
DB 190 SGQQQQRVAIRALRIKEPRIILFGDEPTGAVNDISKILQPFVEYNRDKGTIVLVTHNE 249
QY 192 LFRESADTIKL 203
DB 250 KIVELAKRVIKI 261

RESULT 10
Y796.METJ
ID Y796.METJA STANDARD; PRT; 235 AA.
AC Q58206;
DT 01-NOV-1997 (Rel. 35; Created)
DT 01-NOV-1997 (Rel. 35; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE ABC transporter ATP-binding protein MJ0796.
GN MJ0796.
OS Methanococcus jannaschii.

```

```

106 KHKKWPIPLQVHLHVGLESTEE-----KPVFELSGGCGQORTALARVLLKNPRILLADEPTG 174
121 RRR-----RALECCCKWAELEERFANHKP-NOLSGGCGQORVAIRALANPPILLADEPTG 174
161 ALDLTNSLVIAELRALADK-GATVVVATHSPFLRESADTIKL 203
175 ALDSKTGKIMQLLKKINEEDGKTVVVVTHDINVARFGERIIVL 218

RESULT 11
ID _LOLD ECOLI
AC P75957; Q9R7N6; STANDARD; PRT; 233 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lipoprotein releasing system ATP-binding protein lold.
GN LOLD OR Bll17.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1685;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474 (1997).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 7.18-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RN DNA Res. 3:137-155 (1996).
RL [3]
RP SEQUENCE OF 1-11, AND CHARACTERIZATION.
RX MEDLINE=92045870; PubMed=10783239;
RA Yakushi T., Masuda K., Narita S., Matsuyama S., Tokuda H.;
RT "A new ABC transporter mediating the detachment of lipid-modified
RT proteins from membranes.";
RN Nat. Cell Biol. 2:212-218 (2000).
CC -1- FUNCTION: Part of an ATP-dependent transport system responsible
CC for the release of lipoproteins targeted to the outer membrane
CC from the inner membrane. Such a release is dependent of the
CC sorting-signal (absence of an Asp at position 2 of the mature
CC lipoprotein) and of IolA.
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated.
CC -1- SIMILARITY: Belongs to the ABC transporter family. Lold subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement)
CC or send an email to license@sib-sib.ch.
CC -----
CC EMBL; AE000212; AAC74201.1; ALT_INIT.
CC EMBL; D90747; BAA35937.1; ALT_INIT.
CC HSP; Q58663; IG6H.
CC EcoGene; EG13440; lold.
CC InterPro; IPR003593; AAA ATPase.
CC -----

```


DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW Hypothetical protein; Transport; ATP-binding; Complete proteome.
FT NP_BIND 38 45 ATP (POTENTIAL).
SQ SEQUENCE 224 AA; 25263 MW; CF0B72742D24921E CRC64;

Query Match 30.3%; Score 309; DB 1; Length 224;
Best Local Similarity 34.3%; Pred. No. 4.4e-19;
Matches 73; Conservative 40; Mismatches 90; Indels 10; Gaps 2;

QY 1 MIEINDLKSKFG---VAILWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPSSG 56
DB 1 MIEAKVWKIYKGEAKTIALKNLNKIEGFEVWINGPSGCGKSTLLNCLALDTPTKG 60

QY 57 QILVEDVLLKSTRKQRLRYKNTVGYLFQDYALIPDRTVKFNQLQAV-----EKHKWP 110
DB 61 EYVYKGRATSSMSENERAIFRKISGFIQFOFHILKTLTALENVFLPMFLDERDKSYRK 120

QY 111 EIPQVHAIVGLSEPEERKVPFELSGGEQRTALARVLKPNRIILADEPTGALDITNSLV 170
DB 121 RAKLLENVGLDRLNHYPHQLSGGQQQORVAIRALANPKIIFADEPTGNLSKSMV 180

QY 171 TEALRALADKCATVVVATHSPFPRESADTIKL 203
DB 181 MSILKGLNEKGITIMVTHQELTKYASKIIL 213

RESULT 13
METN_VIBCH
ID METN_VIBCH STANDARD; PRT; 344 AA.
AC Q9KJ5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable D-methionine transport ATP-binding protein metN.
GN METN OR VC0907.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.J., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae";
RL Nature 406:477-483 (2000).
CC -!- FUNCTION: Part of the binding-protein-dependent transport system
CC metNIQ for D-methionine. Probably responsible for energy coupling
CC to the transport system (by similarity).
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (Potential).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE004174; AAF94069.1; -.
CC PIR; C82286; C82266.
CC TIGR; VC0907; -.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003439; ABC transporter.
CC

DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport; Complete Proteome.
FT NP_BIND 42 49 ATP (BY SIMILARITY).
SQ SEQUENCE 233 AA; 25438 MW; IC305476182F1EFD CRC64;

Query Match 30.4%; Score 303.5; DB 1; Length 233;
Best Local Similarity 42.4%; Pred. No. 4.1e-19;
Matches 72; Conservative 31; Mismatches 60; Indels 7; Gaps 3;

QY 27 GNTALTGASGSGKSTLLNCLGTLDPSSGQILVEDVLLKSTRKQRLRYKNTVGYLFQ 86
DB 35 GEMMAIVGSSGSGKSTLLHLLGGDTPTSNGVIFNGQFMSLSSAAAELENKLGFIYQ 94

QY 87 DYALIPDRTVKFN--LQLAVERKHWPETP-----OVLHAGVGLSEPEERKVPFELSGGEQRT 140
DB 95 PHILIPDPTALENVAMPILLIKKPAEINLSRALEMLKAVGLDHRANRPPSELSSGGERQV 154

QY 141 ALARVLKPNRIILADEPTGALDITNSLSVIEALRALAD-KGATVVVATH 189
DB 155 ALARALVNNPRVLADEPTGNLDARNADSIQFLGELNRLQGTAFVLVTH 204

RESULT 12
YF08_METJA
ID YF08_METJA STANDARD; PRT; 224 AA.
AC Q58903;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein MJ1508.
GN MJ1508.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weisman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii";
RL Science 273:1058-1073 (1996).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67591; AAB99520.1; -.
CC PIR; C64488; C64488.
CC TIGR; MJ1508; -.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003439; ABC_transporter.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; PD000006; ABC_transporter; 1.
CC SMART; SM00382; AAA; 1.
CC

Azevedo V., Bertero M.G., Bessieres P., Bolorin A., Borcherst S.,
RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Broutillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaeser P., Goffeau A., Golightly E.J.J., Grandi G.,
RA Guiseppi G., Guy B.J., Hega K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karanata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinou S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Madigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Porretele D., Porwollik S., Prescott A.M.,
RA Pressecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sakiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzengatter T.,
RA Winters P., Wijat P., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
NT Nature 390:249-256 (1997).
RN [5]
RP SEQUENCE OF 66-247 FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=93360813; PubMed=835609;
RA Cosmina P., Rodriguez F., de Ferla F., Grandi G., Perego M.,
RA Venena G., van Sinderen D.;
RT "Sequence and analysis of the genetic locus responsible for surfactin
RT synthesis in Bacillus subtilis";
ML Microbiol. 8:821-831(1993).
CC CC - FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM FOR AN AMINO ACID. PROBABLY RESPONSIBLE FOR ENERGY COUPLING
CC TO THE TRANSPORT SYSTEM.
CC - SIMILARITY: Belongs to the ABC transporter family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement.(See http://www.isb-sib.ch/announce/
CC or send an email to licensee@lsb-sib.ch).

DR EMBL; X77636; CAAS4726.1; -
DR EMBL; L17438; AAC36828.1; -
DR EMBL; D50453; BAA08992.1; -
DR EMBL; Z99105; CAB12153.1; -
DR EMBL; X70356; -; NOT_ANNOTATED_CDS.
DR PIR; C69761; C69761.
DR Subtilist; BG11185; yckI.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF000005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SMC0382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
FW ATP-binding; Transport; Complete proteome.
KW NP_BIND 34 41 ATP (POTENTIAL).
FT FT BIND 72 93
FT CONFLICT 84 93
FT CONFLICT 84 84 SGVSPV (IN REF. 2)).
FT CONFLICT 247 AA; 27788 MW; D08F42F2B3490A7D CRC64;
SQ SEQUENCE

```

Query Match      29.7%; Score 303; DB 1; Length 247;
Best Local Similarity 37.7%; Pred. No. 1.6e-18;
Matches 84; Conservative 35; Mismatches 70; Indels 34; Gaps 8;

Qy 1 MIEINDLK3FGVRIWQGLSHKFLPGTMTALTGASGSGKSTLNCGLTLDKPSGQILV 60
Db 1 MLTVKGLNKSFGENEILKKIDMKIEKVIAILGPGSGKSTLNCGLNALEIENRGELAF 60

Qy 61 ED-----VDLKLSTRKQRLYRKNTVGYLFQDYALIPDRVKNLQ---LAVEK 106
Db 61 DDFSIDFSKKVQADILKLR-----RKSPM--VFQAVHLFPHRTALENVMSGPVQVQK 111

Qy 107 HKWEIP-----QVLHVAVLGSFBEKPVFELSGGQRTALARVLLKNPRIILADEPTGAL 162
Db 112 RNKEVRKEAQLDLKVLKDKMDLYFFQLSGGQQRVGARALAIQPELMFLDEPTSL 171

Qy 163 DLTNSELVIALRA---LADKATVVVATHSPULF-RESADTII 201
Db 172 D---PELVGEVLKVIKDLANEGTMVVVTHEIRFAQEVADVI 211

RESULT 15
YXEO_BACSU STANDARD; PRT; 249 AA.
ID YXEO_BACSU STANDARD; PRT; 249 AA.
AC PS4954;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable amino-acid ABC transporter ATP-binding protein yxeO.
GN YXEO OR JP9G OR BSU39480.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC141;
RA Yeshida K.-I., Fujimura M., Yanai N., Fujita Y.;
RT "Cloning and sequencing of a 23-kb region of the Bacillus subtilis
RL genome between the iol and hut operons.";
RN DNA Res. 2:295-301(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enlart K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Gim S.-Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haesch J., Harwood C.R., Renaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigic C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler T., Weitzneger T.,
RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;

```

```

RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM FOR AN AMINO ACID. PROBABLY RESPONSIBLE FOR ENERGY COUPLING
CC TO THE TRANSPORT SYSTEM.
CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D45912; BAA08331.1; -
CC EMBL; Z99124; CAB15984.1; -
CC PIR; A70076; A70076.
CC SubtilList; BG11891; yxeO.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transporter.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; PD000006; ABC_transporter; 1.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
CC PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
CC KW Hypothetical protein; ATP-binding; Transport; Membrane;
KW Complete proteome.
FT NP_BIND 34 41 ATP (POTENTIAL).
SQ SEQUENCE 249 AA; 27742 MW; A63886EDE69AB80B CRC64;
Query Match      29.7%; Score 302.5; DB 1; Length 249;
Best Local Similarity 34.6%; Pred. No. 1.7e-16;
Matches 74; Conservative 42; Mismatches 81; Indels 17; Gaps 5;

Qy 1 MIEINDLK3FGVRIWQGLSHKFLPGTMTALTGASGSGKSTLNCGLTLDKPSGQILV 60
Db 1 MLTVKIRKAFKDLWLDGIDLVKRGVVAIIIGPSGSGKSTLNCGLNLLERPDQGLIEI 60

Qy 61 EDVLD-LKLSTRKQRLYRKNTVGYLFQDYALIPDRVKNL--QLAVEKHWEIP--- 113
Db 61 GEAKLNAEKFTKEAHLRQQTAMVFQYNLFKNKTALQNTALIVAHK----PRDEA 116

Qy 114 -----QVLHVAVLGSFBEKPVFELSGGQRTALARVLLKNPRIILADEPTGALDTNSE 168
Db 117 KRIGMEILKQVGLSHKADSYFITMSGGQQRIGIARALAVNPHAILLDEPTSLDPBLVT 176

Qy 169 LVIEALRALADKATVVVATHSPULF-RESADTII 201
Db 177 GVLQVKSIAEKQTMIIIVTHEMAFAKEVADQVI 210

```

Search completed: July 27, 2004, 16:35:28
Job time : 14 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 16:32:35 ; Search time 38 Seconds
(without alignments)
1685.531 Million cell updates/sec

Title: US-09-868-338-9
Perfect score: 1019
Sequence: 1 MEINDLKSFGVRLWQGL.....VVVATHSLFRESADTIKL 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL 25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	411	40.3	211	16 Q93ME0	Q93me0 clostridium
2	408	40.0	213	16 Q97RU9	Q97ru9 streptococc
3	407	39.9	212	16 Q892J8	Q892j8 clostridium
4	406	39.8	213	2 Q92HB1	Q92hb1 streptococc
5	405	39.7	213	16 Q8DQ3	Q8dqm3 streptococc
6	399	39.2	213	16 Q97NN3	Q97nn3 streptococc
7	399	39.2	213	16 Q8DND2	Q8dnd2 streptococc
8	386	37.9	213	16 Q97T44	Q97t44 streptococc
9	385	37.8	213	16 Q8DR17	Q8dr17 streptococc
10	382	37.5	207	2 Q9L550	Q9l550 lactococcus
11	380.5	37.3	210	16 Q97NR1	Q97nr1 streptococc
12	380.5	37.3	210	16 Q8DNF2	Q8dnf2 streptococc
13	375	36.8	228	16 Q9K767	Q9k767 bacillus ha
14	373	36.6	211	16 Q9CJ13	Q9cj13 lactococcus
15	371.5	36.5	248	16 Q93H50	Q93h50 streptomyce
16	366	35.9	208	16 Q9AC69	Q9ac69 staphylococ

17	361.5	35.5	226	17	O28803	archaeoglob
18	361	35.4	648	16	Q83LR7	shigella fl
19	360	35.3	232	17	Q973J6	sulfolobus
20	358	35.1	381	16	Q8YT2	brucella me
21	357.5	35.1	224	16	Q8DY9	fusobacteri
22	356.5	35.0	246	16	Q9LOJ9	streptomyce
23	356	34.9	233	16	Q8POU2	streptococc
24	355	34.8	213	16	Q9KG34	bacillus ha
25	354.5	34.8	228	17	O29244	archaeoglob
26	353	34.6	227	16	Q9K6S5	bacillus ha
27	352	34.5	233	16	Q992F0	streptococc
28	351.5	34.5	216	16	Q8LHP0	bacillus ce
29	351	34.4	232	16	Q97IV5	clostridium
30	351	34.4	233	16	Q8K7B3	streptococc
31	351	34.4	237	16	Q878P3	streptococc
32	350.5	34.4	300	16	Q9PR26	ureaplasma
33	349.5	34.3	255	16	Q97IX2	clostridium
34	349	34.2	246	17	Q9UX77	sulfolobus
35	348.5	34.2	265	17	Q8PTP8	methanosarc
36	348	34.2	233	16	Q8XME3	clostridium
37	348	34.2	250	16	Q8ESM1	streptococc
38	348	34.2	250	16	Q8DX1	streptococc
39	346	34.0	224	16	Q99XU3	streptococc
40	346	34.0	231	17	O28456	archaeoglob
41	345.5	33.9	217	16	Q9PM14	campylobact
42	345.5	33.9	227	16	Q894W5	clostridium
43	345.5	33.9	664	16	Q88UI5	lactobacilli
44	344.5	33.8	227	16	Q9CNI3	pasteurella
45	344	33.8	207	16	Q8Y3Q9	listeria mo

ALIGNMENTS

RESULT 1

Q93ME0	PRELIMINARY;	PRT;	211 AA.
ID Q93ME0			
AC Q93ME0			
DT 01-DEC-2001	(TRENBLREL. 19, Created)		
DT 01-DEC-2001	(TRENBLREL. 19, Last sequence update)		
DT 01-OCT-2003	(TRENBLREL. 25, Last annotation update)		
DE	Probable ABC transporter.		
GN PCP07.			
OS	Clostridium perfringens.		
OG	Plasmid pCP13.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_taxid=1502;		
[1]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=13 / Type A;		
RX	MEDLINE=21664373; PubMed=11792842;		
RA	Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,		
RA	Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,		
RT	Complete genome sequence of Clostridium perfringens, an anaerobic		
RT	flesh-eater.		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).		
DR	EMBL; AP003515; BAB62445.1;		
DR	GO; GO:0046821; C:extrachromosomal DNA; IEA.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.		
DR	GO; GO:0000166; F:nucleotide binding; IEA.		
DR	InterPro; IPR003593; AAA ATPase.		
DR	InterPro; IPR003439; ABC transporter.		
DR	Pfam; PF00005; ABC tran; 1.		
DR	ProDom; PD000006; ABC transporter; 1.		
DR	SMART; SM00382; AAA; 1.		
DR	PROSITE; PS00211; ABC TRANSPORTER_1; 1.		
DR	PROSITE; PS00893; ABC TRANSPORTER_2; 1.		
KW	Plasmid; Complete proteome.		
SQ	SEQUENCE 211 AA; 23971 MW; D490252EB93B5724 CRC64;		


```
Q9ZHB1
ID Q9ZHB1 PRELIMINARY; PRT; 213 AA.
AC Q9ZHB1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RC STRAIN=GS4;
RC MEDLINE=99061139; PubMed=9846742;
RX Massidda O., Anderluzzi D., Friedli L., Feger G.;
RA "unconventional organization of the division and cell wall gene
RT cluster of Streptococcus pneumoniae.";
RL Microbiology 144:3069-3078(1998).
CC -f- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF068902; AAC95448.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transporter; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR KW Hypothetical protein; ATP-binding; Transporter.
SQ SEQUENCE 213 AA; 23962 MW; 2939F15CB0044F4 CRC64;

Query Match 39.8%; Score 406; DB 2; Length 213;
Best Local Similarity 41.9%; Pred. No. 2,le-25;
Matches 88; Conservative 43; Mismatches 69; Indels 10; Gaps 3;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPGTWTALTGASGSKTLNCLGTLDPKSSGQILV 60
DB 1 MIELKNISKFGSRQLFSDTNLHFEGGKIYALIGTSCGKTTLLNMGSLRLEPYDKGIIY 60
QY 61 EDVDLLKSTRKQRLYKNTVGYLFQDYALIPDRTVKFNQLAV-----EKHKWPEIPQ 114
DB 61 DGTSLKDI---KSSVFRDYGLYLFQDFGLIESQTVKENLNGLVGKKLKEKESILMKQ 117
QY 115 VLHVGLESPE-EKPVFELSGGQRTALARVLLKNPRIILADEPTGALDLTNSLIVIEA 173
DB 118 ALNRVNLSDLKQPIFELSGGGAQRVALAKIILKDPPIILADEPTASLDPKNSSELLSI 177

QY 174 LRALAKGATVVVATHSPFPRESADTIKL 203
DB 178 LESLKNPNRTIIATHNPLIWEQDVQVIRV 207

RESULT 5
Q8DQM3
ID Q8DQM3 PRELIMINARY; PRT; 213 AA.
AC Q8DQM3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter ATP-binding protein-sodium transport.
GN ABC-NBD OR SPR0602.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
```

```
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Gerlinger C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008439; AAX99406.1; -.
DR FIR; B97947; B97947.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transporter; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_TRANSPORTER.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR KW ATP-binding; Complete Proteome.
SQ SEQUENCE 213 AA; 23972 MW; C4989CCCB0062965 CRC64;

Query Match 39.7%; Score 405; DB 16; Length 213;
Best Local Similarity 41.9%; Pred. No. 2,6e-25;
Matches 88; Conservative 43; Mismatches 69; Indels 10; Gaps 3;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPGTWTALTGASGSKTLNCLGTLDPKSSGQILV 60
DB 1 MIELKNISKFGSRQLFSDTNLHFEGGKIYALIGTSCGKTTLLNMGSLRLEPYDKGIIY 60
QY 61 EDVDLLKSTRKQRLYKNTVGYLFQDYALIPDRTVKFNQLAV-----EKHKWPEIPQ 114
DB 61 DGTSLKDI---KPSVFRDYGLYLFQDFGLIESQTVKENLNGLVGKKLKEKESILMKQ 117
QY 115 VLHVGLESPE-EKPVFELSGGQRTALARVLLKNPRIILADEPTGALDLTNSLIVIEA 173
DB 118 ALNRVNLSDLKQPIFELSGGGAQRVALAKIILKDPPIILADEPTASLDPKNSSELLSI 177

QY 174 LRALAKGATVVVATHSPFPRESADTIKL 203
DB 178 LESLKNPNRTIIATHNPLIWEQDVQVIRV 207

RESULT 6
Q97NN3
ID Q97NN3 PRELIMINARY; PRT; 213 AA.
AC Q97NN3;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter, ATP-binding protein.
GN SPI987.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
```

RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae";
 RL Science 293:498-506(2001).
 DR EMBL; AB007488; AAK76054.1; -.
 DR PIR; E95232; E95232.
 DR PIR; G98096; G98096.
 DR TIGR; SP1987; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Complete proteome.
 SQ SEQUENCE 213 AA; 23920 MW; F6BE2EA4EE87FDB7 CRC64;
 Query Match 39.2%; Score 399; DB 16; Length 213;
 Best Local Similarity 42.9%; Pred. No. 7.9e-25;
 Matches 90; Conservative 40; Mismatches 70; Indels 10; Gaps 3;
 QY 1 MIEINDLKXFGVRLWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
 Db 1 MIDIQGLEKFKFNDRAIFSGNLKLEKGVYALIGKSGKTTLLNIGLKEIDGGRLV 60
 QY 61 EDVLLKLSRKQRLYKNTVGYLFQDYALIPDRTVKFNQLQAVEKHWPPIQVHLHVG 120
 Db 61 QCKDLKTIPTRE---YFDDQGVYLFQNGLENSIKENLDLGFVGQKISKVERLERQVG 117
 QY 121 -LESFE-----EKPVFELSGEQOQTALARVLKNPRIILADEPTGALDITNSLVIEA 173
 Db 118 ALEKVNLYGLDLEQKIVTLTSGGEAQRVALAKILKNPPLILADEPTAALDPENSEEVMNL 177
 QY 174 LRLADKAGTAVVATHSPFRESADTIKL 203
 Db 178 LVLDKDNRIIIATHNPLVWKNKADEIIDM 207
 RESULT 7
 Q8DND2 PRELIMINARY; PRT; 213 AA.
 ID Q8DND2
 AC Q8DND2
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ABC transporter ATP-binding protein-unknown substrate.
 GN ABC-NBD OR SP1801.
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=171101;
 RN [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Albom W.E. Jr., Arnold J., Blaszczyk L.C., BURGATT S.,
 RA DeHoff B.S., Essem S.T., Fritz L., Fu D.-J., Fuller L.W., Geringer C.,
 RA Gilmour R.J., Glass J.H., Knoch J., Kraft A.R., Lagace R.E.,
 RA LeBlanc D.J., Lee L.N., Leikowitz E.J., Lu J., Matsushima P.,
 RA McAhren S.M., McHenry M., McKeen K., Mundy C.W., Nicas T.I.,
 RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
 RA Glass J.;
 RT "genome of the bacterium Streptococcus pneumoniae strain R6";
 RL J. Bacteriol. 183:5709-5717(2001).
 DR EMBL; AE008544; AAL00604.1; -.
 DR PIR; E95232; E95232.
 DR PIR; G98096; G98096.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_AIPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Complete proteome.
 SQ SEQUENCE 213 AA; 23920 MW; F6BE2EA4EE87FDB7 CRC64;
 Query Match 39.2%; Score 399; DB 16; Length 213;
 Best Local Similarity 42.9%; Pred. No. 7.9e-25;
 Matches 90; Conservative 40; Mismatches 70; Indels 10; Gaps 3;
 QY 1 MIEINDLKXFGVRLWQGLSHKFLPGTMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
 Db 1 MIDIQGLEKFKFNDRAIFSGNLKLEKGVYALIGKSGKTTLLNIGLKEIDGGRLV 60
 QY 61 EDVLLKLSRKQRLYKNTVGYLFQDYALIPDRTVKFNQLQAVEKHWPPIQVHLHVG 120
 Db 61 QCKDLKTIPTRE---YFDDQGVYLFQNGLENSIKENLDLGFVGQKISKVERLERQVG 117
 QY 121 -LESFE-----EKPVFELSGEQOQTALARVLKNPRIILADEPTGALDITNSLVIEA 173
 Db 118 ALEKVNLYGLDLEQKIVTLTSGGEAQRVALAKILKNPPLILADEPTAALDPENSEEVMNL 177
 QY 174 LRLADKAGTAVVATHSPFRESADTIKL 203
 Db 178 LVLDKDNRIIIATHNPLVWKNKADEIIDM 207
 RESULT 8
 Q97T44 PRELIMINARY; PRT; 213 AA.
 ID Q97T44
 AC Q97T44
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Amino acid ABC transporter, ATP-binding protein, putative.
 GN SP0111.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 SEQUENCE FROM N.A.
 RX STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Uutterback T.R., Hansen C.L.,
 RA McDonald L.A., Keldblum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae";
 RL Science 293:498-506(2001).
 DR EMBL; AE007327; AAK74298.1; -.
 DR PIR; A95013; A95013.
 DR TIGR; SP0111; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_AIPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.

DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23640 MW; E35696AEB3677399 CRC64;

Query Match 37.9%; Score 386; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPQMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKTGGKVLNLSLRIDQDLVAIVGKSGSGKSTLLNLLGLIDGYSGRYEI 60
QY 61 EDVDLLKLSRQELRYKNTVGYLFDYVALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIREHISYLFQNFALIDDETVEYNMLALKYVKLPKDKKLKXVEE 120
QY 115 VLHVGLESFEEKPVPFELSGEQOQTALARVLLKPNRIILADEPTGALDTNSELVIEAL 174
DB 121 ILERVGLSATLHQRVSELSGGEQOQRIAVARAILKPSQILADEPTGSLDPENRDLVKFL 180
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEEL 209

RESULT 9
QDR17
ID Q8DR17 PRELIMINARY; PRT; 213 AA.
AC Q8DR17;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ABC transporter ATP-binding protein-aminic acid transport.
GN ABC-NBD OR SPR01000.
OC Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAnren S.M., McEneaney W., McEneaney K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Young Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AB008393; AAK98904.1; -.
DR PIR; D97884; D97884.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0001166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23626 MW; 205696AEB3677388 CRC64;

Query Match 37.8%; Score 385; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPQMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKTGGKVLNLSLRIDQDLVAIVGKSGSGKSTLLNLLGLIDGYSGRYEI 60
QY 61 EDVDLLKLSRQELRYKNTVGYLFDYVALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIREHISYLFQNFALIDDETVEYNMLALKYVKLPKDKKLKXVEE 120
QY 115 VLHVGLESFEEKPVPFELSGEQOQTALARVLLKPNRIILADEPTGALDTNSELVIEAL 174
DB 121 ILERVGLSATLHQRVSELSGGEQOQRIAVARAILKPSQILADEPTGSLDPENRDLVKFL 180
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEEL 209

RESULT 9
QDR17
ID Q8DR17 PRELIMINARY; PRT; 213 AA.
AC Q8DR17;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ABC transporter ATP-binding protein-aminic acid transport.
GN ABC-NBD OR SPR01000.
OC Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAnren S.M., McEneaney W., McEneaney K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Young Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AB008393; AAK98904.1; -.
DR PIR; D97884; D97884.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0001166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 213 AA; 23626 MW; 205696AEB3677388 CRC64;

Query Match 37.8%; Score 385; DB 16; Length 213;
Best Local Similarity 40.2%; Pred. No. 9.2e-24;
Matches 84; Conservative 45; Mismatches 74; Indels 6; Gaps 1;

QY 1 MIEINDLKSGFVRILWQGLSHKFLPQMTALTGASGSGKSTLLNCLGTLDPKSSGQILV 60
DB 1 MIELKNITKTGGKVLNLSLRIDQDLVAIVGKSGSGKSTLLNLLGLIDGYSGRYEI 60
QY 61 EDVDLLKLSRQELRYKNTVGYLFDYVALIPDRTVKFNQLAV-----EKHKWPEIQ 114
DB 61 FGQTNLAVNSAKSQTIIREHISYLFQNFALIDDETVEYNMLALKYVKLPKDKKLKXVEE 120
QY 115 VLHVGLESFEEKPVPFELSGEQOQTALARVLLKPNRIILADEPTGALDTNSELVIEAL 174
DB 121 ILERVGLSATLHQRVSELSGGEQOQRIAVARAILKPSQILADEPTGSLDPENRDLVKFL 180
QY 175 RALADKAGTAVVATHSPFLFRESADTIKL 203
DB 181 LEMNREGKTVIIVTHDAYVAQOCHRIEEL 209

RESULT 9
QDR17
ID Q8DR17 PRELIMINARY; PRT; 213 AA.
AC Q8DR17;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ABC transporter ATP-binding protein-aminic acid transport.
GN ABC-NBD OR SPR01000.
OC Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAnren S.M., McEneaney W., McEneaney K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Young Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AB008393; AAK98904.1; -.
DR PIR; D97884; D97884.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0001166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATP

```

QY 174 LRALADKAGATVVVATHSPFLPRESADTIKL 203
Db 178 LLSLVKENTVIIATHSHVNRVDSTIINL 207

RESULT 11
Q37NR1
ID Q37NR1 PRELIMINARY; PRT; 210 AA.
AC Q37NR1
DT 01-OCT-2001 (TREMELrel. 18, Created)
DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
DE ABC transporter ATP-binding protein-unknown substrate.
DE ABC transporter, ATP-binding protein.
GN SP1957.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lerkowitz E.J., Lu J., Matsushima P.,
RA McAnen S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008541; AAL00576.1; -
DR PIR; C98093; C98093.
DR TIGR; SP1957; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004005; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00092; N6_MTASE; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 210 AA; 23751 MW; EC0373F9B66FE167 CRC64;

Query Match 37.3%; Score 380.5; DB 16; Length 210;
Best Local Similarity 41.0%; Pred. No. 2.5e-23;
Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;

QY 1 MIEINDLKSGVRIWMGLSHKFLPTMTALTGASGSGKSTLLNCIGTLDPKPSGGILV 60
Db 1 MIELQVSKSGFERELFNSMTFAGKVYALIGSSGSGKTLNMNMGKLE-PYDGTIFY 59
QY 61 EDVLLKSLSTRKQRYKNVTGYLYFDYALIPDRVTNFVNLQAVEKKWPEI-----PQ 114
Db 60 RGKD---LANYKSDFFRHEGLYFQNGFLNGSIENKLGIGQKLSRSEQLRQKQ 116
QY 115 VLHVGLESFE-EKVPFELSGEQORTALARVLKNPRIILADEPTGALDUTNSLVEIA 173
Db 117 ALEQVGLVYLLDKRIIFELSGESQORVALAKIILKNPFIILADEPTASIDPATSQLIMEI 176
QY 174 LRALADKAGATVVVATHSPFLPRESADTIKL 203
Db 178 LLSLVKENTVIIATHSHVNRVDSTIINL 207

RESULT 12
Q8DNF2
ID Q8DNF2 PRELIMINARY; PRT; 210 AA.
AC Q8DNF2
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DE ABC transporter ATP-binding protein-unknown substrate.
DE ABC-NBD OR SP1773.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lerkowitz E.J., Lu J., Matsushima P.,
RA McAnen S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008541; AAL00576.1; -
DR PIR; C98093; C98093.
DR TIGR; SP1957; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004005; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00092; N6_MTASE; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 210 AA; 23751 MW; EC0373F9B66FE167 CRC64;

Query Match 37.3%; Score 380.5; DB 16; Length 210;
Best Local Similarity 41.0%; Pred. No. 2.5e-23;
Matches 86; Conservative 41; Mismatches 72; Indels 11; Gaps 4;

QY 1 MIEINDLKSGVRIWMGLSHKFLPTMTALTGASGSGKSTLLNCIGTLDPKPSGGILV 60
Db 1 MIELQVSKSGFERELFNSMTFAGKVYALIGSSGSGKTLNMNMGKLE-PYDGTIFY 59
QY 61 EDVLLKSLSTRKQRYKNVTGYLYFDYALIPDRVTNFVNLQAVEKKWPEI-----PQ 114
Db 60 RGKD---LANYKSDFFRHEGLYFQNGFLNGSIENKLGIGQKLSRSEQLRQKQ 116
QY 115 VLHVGLESFE-EKVPFELSGEQORTALARVLKNPRIILADEPTGALDUTNSLVEIA 173
Db 117 ALEQVGLVYLLDKRIIFELSGESQORVALAKIILKNPFIILADEPTASIDPATSQLIMEI 176
QY 174 LRALADKAGATVVVATHSPFLPRESADTIKL 203
Db 178 LLSLVKENTVIIATHSHVNRVDSTIINL 207

```



```

RC STRAIN=MA-4680 / ATCC 31267 / NCIMS 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=13692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AB070951; BAB9341.1; -.
DR EMBL; AF005035; BAC71346.1; -.
DR GO; GO:0015020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; transport; Complete proteome.
KW ATP-binding; transport; Complete proteome.
SQ SEQUENCE 248 AA, 26276 MW, 43562A602BB357F9 CRC64;

Query Match 36.5%; Score 371.5; DB 16; Length 248;
Best Local similarity 41.0%; Pred. No. 1.7e-22;
Matches 87; Conservative 37; Mismatches 77; Indels 11; Gaps 4

QY 2 IINDLKXSFQ-----VRILMOGLSHKFLPCTMTALTGSSGSKSTLLNCLGLTDXPKSSG 56
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 9 IQRSVSRRYGAGGAVTALDQ-VSLAPPRGTFITAVMPGSGSKSTLLQCAAGLDRTSG 67
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 57 QILVEDVLLKLSTRKQBYLRNTVGYLFQDYALIPDRTVKFN----LQIAVEKHKWPFI 112
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 SVTVGDTETLKLSETKTLTLRRDRIGFVFQAFNLLPSLTAEQNVALPRLIAGRRPKTEV 127
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 POVLHAVGLGSEEPKVPVELSGGQOORTALARVLLKNPRIILADEPTGALDLTNSLVET 172
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 RVLVAQVGLGDRAHRPTMGSGGQOORVALARALITPDVLFQGEPTGALDSQTSREVL 187
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 ALRALAD-KGATVVVATHSPLFRSADTIKL 203
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 LIRGMDVSEGQTVIMVTHDPVAASVADRVNPL 219
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: July 27, 2004, 16:36:25
Job time : 47 secs